Tools and reference standards supporting the engineering and evolution of synthetic biological systems

By

Jason R. Kelly

Bachelor of Science, Biology and Chemical Engineering
Massachusetts Institute of Technology, 2003

Submitted to the Department of Biological Engineering in
Partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering
at the
Massachusetts Institute of Technology

May 2008

Signature of Author: _____
Department of Biological Engineering
May 23, 2008

Certified by: _____
Drew Endy
Assistant Professor of Biological Engineering
Thesis Supervisor

Accepted by: _____
Alan J. Grodzinsky
Professor of Electrical, Mechanical, Biological Engineering
Chairman, Course XX Graduate Program Committee

This doctoral thesis has been examined by a committee of the Biological Engineering Department as follows:

Thesis Committee Chairman: .....

Edward F. DeLong
Professor of Biological Engineering
Professor of Civil and Environmental Engineering

Thesis Supervisor, Committee Member: .....

Drew Endy
Cabot Assistant Professor of Biological Engineering

Thesis Committee Member:.....

Thomas F. Knight, Jr.
Senior Research Scientist, Computer Science & Artificial Intelligence Laboratory

Abstract

Biological engineers have constructed a number of multi-part synthetic biological systems that conduct logical operations on input signals, produce oscillatory output signals, store memory, or produce desired products. However, very few of these genetically-encoded systems worked as originally designed. The typical process of constructing a functional system involves a period of tuning the system properties to find a functional variant. This tuning process has been optimized and applied with great success to the engineering of individual biological parts by directed evolution. For instance, researchers developing improved enzymes, transcriptional promoters, and fluorescent proteins have generated large libraries of variants and screened these libraries to find individual mutants that met desired performance specifications. In this thesis, I address some of the bottlenecks preventing the application of directed evolution to more complex devices and systems. First, I describe an input / output screening plasmid that was designed to enable screening of higher-order genetic devices based on the equilibrium response of the device. This plasmid includes two fluorescent reporters and an inducible promoter to enable screening of device libraries across a range of inputs. Second, I describe measurement kits and reference standards designed to improve the characterization of promoter and RBS parts that are used as input substrates for device evolution. By using the kits, researchers are able to report promoter and RBS activities in standard units (Standard Promoter Units, SPUs, and Standard RBS Units, SRUs) enabling the growth of a collection of well-characterized parts to draw on for assembling device variants. Finally, I describe a new microfluidic device, the Sortostat, that integrates a cell sorting chamber with a previously published microscope-mounted microfluidic chemostat. Researchers can use the Sortostat to apply morphological, time-varying, or other complex selective pressures to cells in continuous culture.

Acknowledgements

To my wife-to-be Hillary, thanks for being a perfect balance to me. I would be completely off the rails if you weren't around. To my parents and my sister Amanda, thanks for always supporting me 100% and helping me remember what really matters.

I'd like to thank Tom Knight for showing me what a real hacker looks like. It's such an unbelievably rare phenotype that it's startling to actually meet one. Thanks to Ed Delong for providing helpful comments throughout my thesis, and to Frederick Balagadde for teaching me to design and debug microfluidic chips.

Thanks to the OpenWetWare posse (Sri, Ilya, Reshma, Barry, Austin) for basically starting up a small business on the side of their graduate projects. Working as a team to boot up OWW was one of the best parts of my graduate experience (making E.coli smell like bananas was pretty sweet too) and it's in large part due to working with such a thoughtful, talented group.

I have to use "I" throughout this thesis because it's apparently some convention, but it is ridiculous to suggest that the work here is mine alone. I've been privileged to work with no less than seven MIT undergraduate students: Josh Michener, Kelly Chang, Bryan Hernandez, Andrzej Wojcieszynski, Karen Wong, Nishant Baht, and Adam Rubin. Without their help I literally wouldn't have a thesis.

I've been blessed with so many amazing friends that I really can't list them. Special thanks though to Barry, Amy, and Victor for tolerating me as a housemate for 4 years - the Buddha Lounge will live forever.

I have wanted to hack biological systems for a really long time. As a freshman at MIT I rushed to the lab and spent the following four summers realizing that even at the best engineering school in the world they didn't know how to do it. Drew, thanks for changing my mind.

Table of Contents

Chapter 1. Introduction 12

Chapter 2. pSB1A10, A PoPS Input/Output Characterization and Screening Plasmid. 18

2.1 Summary 18

2.2 Introduction 19

2.3 Results and Discussion 24
- 2.3.1 pSB1A10 design & construction 24
- 2.3.2 Characterization of six BioBrick Transcription Terminators 29
- 2.3.3 Characterization of nine BioBrick Inverters 32
- 2.3.4 Repair and Reuse of a Tet repressor-based inverter 37
- 2.3.5 Testing inverters in series, BBa_Q04401 and BBa_Q04740 43
- 2.3.6 pSB3K10, a low copy, simplified PoPS input / output characterization and screening plasmid 46

2.4 Conclusions 47

Chapter 3. Measurement kits and reference standards for BioBrick promoters and ribosome binding sites. 49

3.1 Summary 49
- 3.1.1 Background 49
- 3.1.2 Results 50
- 3.1.3 Conclusions 50

3.2 Introduction 51

3.3 Results 60
- 3.3.1 Measurement kit design & construction 60
- 3.3.2 Definitions and models for promoter and RBS activity 63
- 3.3.3 Demonstration of Measurement kit utility 69
- 3.3.4 Variability due to measurement procedures 71
- 3.3.5 Inter-laboratory variation 73

3.4 Discussion 75
- 3.4.1 Suggested improvements 76
- 3.4.2 Standard promoter and RBS definition 79
- 3.4.3 Measurement procedures 80
- 3.4.4 Absolute and relative units 81
- 3.4.5 Distribution, use, and improvement of kits 82

3.5 Conclusions 83

Chapter 4. Sortostat: An integrated microchemostat and optical cell sorting system 84

4.1 Summary 84

4.2 Introduction 85

4.3 Results and Discussion 88
4.3.1 Design and operation of the microchemostat 88
4.3.2 Design and Fabrication of the Sortostat microfluidic chip 92
4.3.3 Sortostat automated platform 95
4.3.4 Operating the Sortostat 96
4.3.5 Models of Sortostat function 99
4.3.6 Proof of principle experiments 100
4.3.7 Demonstration of chemostat operation and verification of expected statistical distributions 101
4.3.8 Demonstration of time-varying selective pressure in the Sortostat 106
4.3.9 Rescue of a slower growing subpopulation in the Sortostat 109
4.3.10 Improvements for future versions of the Sortostat 112

4.4 Conclusions 114

Chapter 5. Future Work 116

5.1 Introduction 116

5.2 Scaffold for tuning inverters by inserting libraries of parts. 118

5.3 Part specification for BioBrick promoters 120
5.3.1 “Fauxmoters” 122

5.4 Models for describing the operation of the Sortostat 123

Appendix A. Materials and Methods for Chapter 1 127

A.1 Bacterial Strains, Media, and Chemicals 127

A.2 Characterization of BioBrick Inverters and Terminators 127

A.3 Construction of inverter libraries 128

A.4 Inverters in series 129

Appendix B. Materials and Methods for Chapter 2 131

B.1 Strains and Media 131

B.2 Kit contents 131
B.3 Construction of test constructs 132
B.4 Assay of promoter-RBS collection 133
B.5 Assay of different measurement conditions 134
B.6 Assay of inter-laboratory variability 136

Appendix C. Materials and Methods for Chapter 3 138

C.1 Details of PDMS chip design and fabrication 138
C.2 Previous designs of the Sortostat 139
C.3 Details of Sortostat automated platform 141
C.4 Comparison of the growth rate of CFP expressing and YFP expressing cells. 143
C.5 Image processing 144
C.6 Thin channels for improved image processing 144
C.7 Common failure modes 145

Appendix D. Supplementary materials for Chapter 2 147

D.1 Modeling 147
D.1.1 Derivation of the relationships between absolute promoter or RBS activity and GFP synthesis rate or GFP concentration 147
D.1.2 Derivation of promoter activity in SPUs as a function of GFP synthesis rate 148
D.1.3 Derivation of promoter activity in SPUs as a function of GFP concentration 149
D.1.4 Derivation of RBS activity in SRUs as a function of GFP synthesis rate 150
D.1.5 Derivation of RBS activity in SRUs as a function of GFP concentration 152
D.2 Alternate designs for the RBS and promoter measurement kits 153
D.3 Transcription start site prediction 154
D.4 Calculating the relationship between PoPS and SPUs 154
D.5 Calculating the relationship between protein production rate and SRUs 156

Chapter 6. References 163

List of Figures

Figure 1-1 Directed evolution mapped onto an abstraction hierarchy for engineering biological systems. 15

Figure 2-1 Schematic of the PoPS Input/ Output characterization and screening plasmid (pSB1A10) 21

Figure 2-2 Examples of PoPS Input / Output devices 23

Figure 2-3 RNAseE sites were designed to insulate the reporter mRNA from the test part or device mRNA. 27

Figure 2-4 RNAseE sites help to insulate reporter mRNA stability 28

Figure 2-5 Characterization of the termination efficiency of six BioBrick terminators 31

Figure 2-6 Characterization of BBa_Q04740 35

Figure 2-7 Summary of triplicate experiments characterizing the inverter BBa_Q04740. ... 36

Figure 2-8 Two rounds of screening on a FACS machine of an inverter library generated via mutatgenic PCR of the non-functional inverter, BBa_Q04400. 39

Figure 2-9 Characterization of BBa_Q04401, a successfully repaired inverter. 40

Figure 2-10 Successful reuse of inverter BBa_Q04401 by the MIT iGEM team in an engineered system for controlling bacterial odor. 42

Figure 2-11 Schematic of pSB3K10. 47

Figure 3-1 The British Standard Ohm resistor is an early example of a reference standard. 54

Figure 3-2 Overview of using the measurement kit 56

Figure 3-3 Instructions for inserting a promoter into the promoter measurement kit and measuring the promoter activity in Standard Promoter Units (SPUs). 58

Figure 3-4 Measurement of the activity of a test set of (A) promoters and (B) RBSs using the measurement kits. 70

Figure 3-5 Measurement of the activity of a set of four promoters using four different measurement procedures 72

Figure 3-6 Comparison of promoter activities measured by researchers in seven independent laboratories using the measurement kit. 75

Figure 4-1 A schematic of chemostat and a picture of the microchemostat. 86

Figure 4-2 Cutaway view of a “push down” PDMS microfluidic valve 89

Figure 4-3 Microchemostat image with channels colored by food dye. 90

Figure 4-4 Two modes of operation of the microchemostat. 91
Figure 4-5 Final design for the Sortostat microfluidic chip. 94
Figure 4-6 Addition of the sorting chamber to the microchemostat design 95
Figure 4-7 Schematic of the Sortostat automated platform. 96
Figure 4-8 Demonstration of operating the Sortostat as a normal chemostat. 101
Figure 4-9 Demonstration of operating the Sortostat as a normal chemostat. 102
Figure 4-10 Comparison of model distribution to experimental results for the number of cells captured in the sorting chamber. 104
Figure 4-11 Theoretical joint probability distribution of the total number of cells in the sorting chamber and the number of CFP expressing cells in the sorting chamber agrees with experimental data. 105
Figure 4-12 Demonstration of time-varying selection using the Sortostat. 108
Figure 4-13 Demonstration of the rescue of a slower growing subpopulation of cells using the Sortostat. 111
Figure 4-14 Demonstration of the rescue of a slower growing subpopulation of cells using the Sortostat. 112
Figure 5-1 Schematic of the scaffold for tuning tetR-based inverters. 119
Figure 5-2 Comparison of Sortostat model with experimental results. 125
Figure A-1 Relationship between GFP and RFP fluorescence measured by the same procedure used to characterize the inverters (Chapter A.2). 130
Figure C-1 An example of a common fabrication error: valves that don't completely seal. 139
Figure C-2 Schematic of the six reactor design of the Sortostat. 141
Figure C-3 Comparison of the growth curves of the CFP expressing cells (blue) and YFP expressing cells (red) used in all experiments with the Sortostat. 143
Figure C-4 Example of image processing. 144
Figure C-5 Micrograph of 3 μm high channels (arrows) that are used to bring the cells in a microchannel into a narrowed plane of focus. Image courtesy of Frederick Balagadde [83]. 144
Figure D-1 Example of typical growth curves. 157
Figure D-2 Example of GFP synthesis rate time series 158
Figure D-3 Instructions for inserting an RBS into the RBS measurement kit and measuring the promoter activity in Standard RBS Units (SRUs). 159

List of Tables

Table 2-1 Mean termination efficiency of six BioBrick terminators characterized with pSB1A10. 30

Table 2-2 Results of characterization of nine BioBrick inverters. 34

Table 2-3 Listing of inverters previously coupled in synthetic biological systems. 45

Table 3-1 Components of promoter measurement kit 57

Table D-1 Components of RBS measurement kit 161

Table D-2 Listing of the flow cytometer equipment used at the seven laboratories that participated in the inter-laboratory variability study. 162

Table D-3 Growth rates of cells containing the 4 promoter test constructs that were used in the inter-laboratory study measured relative to the reference standard. 162

Chapter 1. Introduction

"There are only two ways we know of to make extremely complicated things. One is by engineering, and the other is evolution." -Danny Hillis [1]

There is a third approach for construction of complicated things that combines forward engineering and evolution. This hybrid approach requires that the substrate be both designable and evolvable. The design and construction of engineered biological systems may be amenable to such an approach. Here, I outline a framework for constructing engineered biological systems that combines elements of forward engineering with evolution in order to increase the likelihood of producing functional systems.

Directed evolution

Directed evolution has been used extensively by biological engineers in developing improved enzymes [2], transcriptional promoters [3], and fluorescent proteins [4]. Additionally, directed evolution has been applied in other fields such as electrical engineering [5] and computer science [6] where algorithms and circuit designs are made to replicate, mutate, and compete under human defined selective pressures. The process of directed evolution can be divided into three general components: (1) inputs, (2) assembly, and (3) screening.

Inputs are the substrate materials that are used to generate libraries of variants of a target system. When evolving biological parts the most common inputs are nucleotides that are assembled to form libraries of DNA sequences. In computer science or electrical engineering the inputs might be logical functions assembled to form executable code [7]. The features of a good input substrate include: physical composability such that the input can be physically combined with other inputs, functional composability such that a collection of inputs assembled together will be expected to produce systems with some function, and well characterized inputs so that libraries can be constructed from the sets of inputs that are most likely to produce the desired system function. Nucleotides provide an excellent input substrate for the directed evolution of engineered biological parts since they can be readily assembled into larger DNA molecules that reliably function within cells. Additionally, the physical properties of nucleotides are well characterized and libraries can be generated that are biased for particular functionality such as forming hairpins [8], having a low melting temperature, or encoding a particular subset of amino acids [9].

High-throughput assembly of the input substrate to construct large libraries of variants of a target system is essential to generate the diversity necessary to drive directed evolution. When evolving biological systems, common approaches for generating diversity include mutagenic PCR or *de novo* DNA synthesis. *De novo* synthesis in particular is very well suited to assembling large, targeted libraries based on user specifications.

Finally, high-throughput screens or selections are needed in order to isolate functional mutants from large libraries of variants. FACS machines, automated 96 well plate-based

assays, and colony-picking robots can be used to rapidly screen libraries of biological systems. However, it is essential that the desired system functionality has a phenotype that can be detected by one of these automated platforms. This requirement for a detectable phenotype becomes more challenging as the complexity of our biological systems increases.

Abstraction

As biological engineers, we can make use of abstraction in order to simplify the design and construction of engineered biological systems. A previously described 'abstraction hierarchy' for standard biological parts was designed to support the rational engineering of biological systems by managing complexity [10]. This hierarchy consists of four layers: DNA, parts, devices, and systems. 'DNA' is genetic material, 'parts' are basic biological functions composed of DNA, 'devices' are any combination of parts that perform a human-defined function, and 'systems' are any combination of devices. I have mapped the three components of the directed evolution process (inputs, assembly, and screening) onto the abstraction hierarchy (Figure 1-1) to consider how we might best make use of standard parts and abstraction to apply directed evolution to more complicated engineered biological systems.

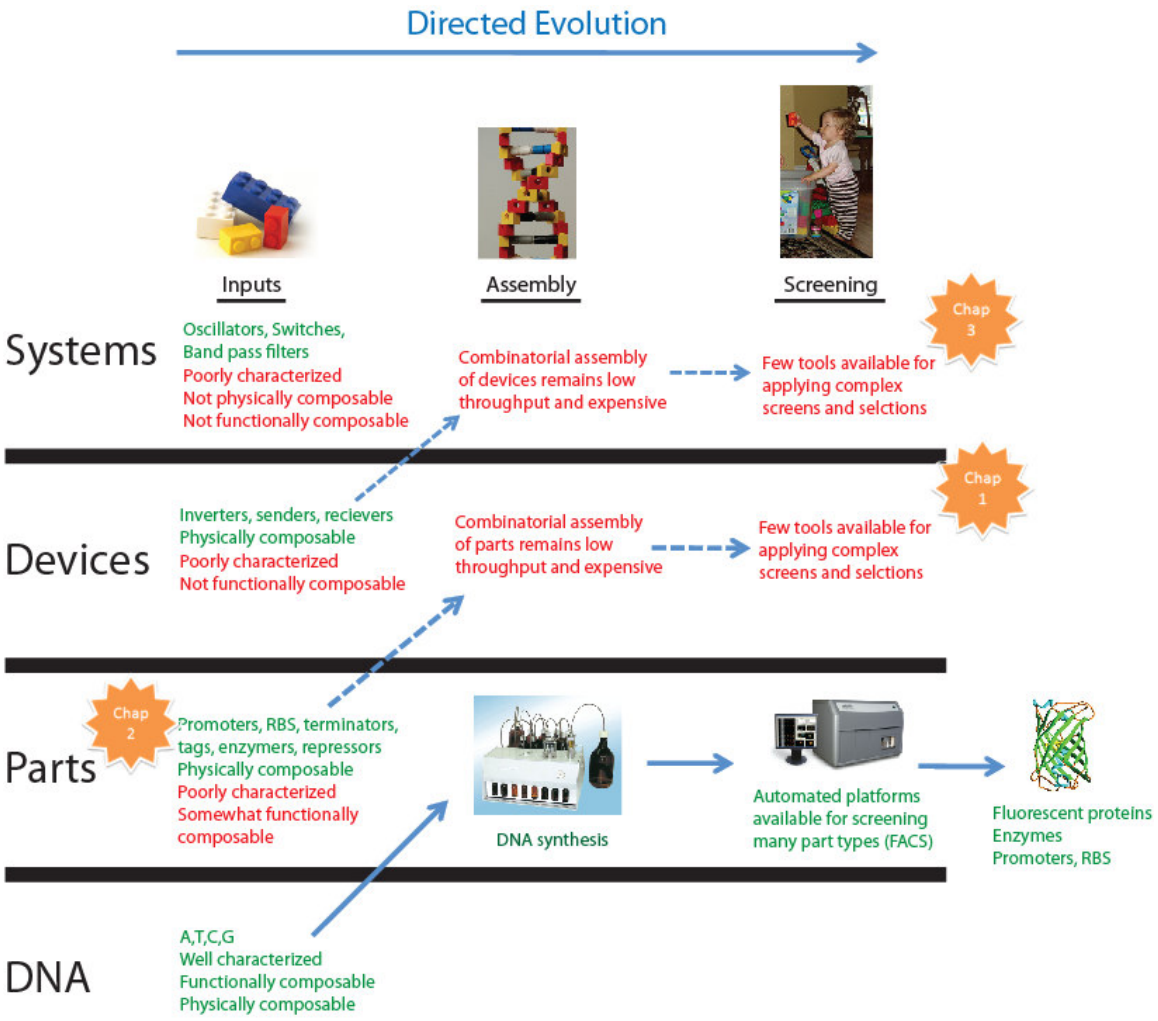


Figure 1-1 Directed evolution mapped onto an abstraction hierarchy for engineering biological systems.

The solid blue arrow represents the successful application of directed evolution of parts. The inputs to the process are DNA nucleotides, these nucleotides can be assembled into enormous libraries in a cheap, automated process (*de novo* DNA synthesis), and high-throughput screens are available for many part functions. A canonical example of the success of this pathway is the directed evolution of new, improved fluorescent protein variants that appear on a nearly monthly basis. Applying directed evolution for the production of devices or systems (dotted blue arrows) remains challenging as there are bottlenecks in all three steps of the directed evolution process: the inputs, the assembly of the inputs into libraries, and the screening of libraries for functional mutants. I have highlighted the area of focus of each of my thesis chapters with orange stars.

Biological engineers might like to apply directed evolution to the construction of parts, devices, or systems that accomplish desired functions. Unfortunately, to date there has only been significant success in the directed evolution of biological parts, though this success has had tremendous impact in both the academia and industry [11, 12]. I have traced a path through the abstraction hierarchy describing the inputs, available technology for assembly, and mechanisms for screening that allow a smooth transition from input nucleotides to evolved parts with desired function such as fluorescent proteins [4] or enzymes [2].

When we try to move up the hierarchy in order to evolve devices such as an inverter [13] we quickly run into bottlenecks in all three steps of the directed evolution process. The input substrate is standard parts such as promoters, ribosome binding sites (RBSs), terminators, and protein coding regions. These parts are often not well defined, nor are they always functionally composable. As an example, an RBS characterized in one context might have a different activity or no activity at all when assembled into a different context [14]. The assembly stage in the directed evolution of devices presents a significant bottleneck as researchers are unable to create large libraries of devices composed of parts. There is not yet the equivalent of the DNA synthesizer for high-throughput assembly of parts, although several techniques begin to address this challenge [15-17]. Finally there exist only a few screening or selection platforms capable of selecting for the more complicated functions associated with devices [18, 19]. For instance, many devices have complicated input / output functions or have an output that varies over time. Moving further up the abstraction hierarchy to evolve systems exacerbates these problems, as the input substrate (devices) is

even less well characterized than parts and harder to functionally combine due to challenges like signal matching between devices [13].

In this thesis I address some of the bottlenecks preventing the application of directed evolution to devices and systems. In Chapter 1, I describe a plasmid, pSB1A10, that was designed to enable screening of devices based on the input / output curve of the device. Additionally, I describe future work constructing a plasmid scaffold that enables easier assembly of parts into inverter libraries. In Chapter 2, I describe measurement kits and reference standards designed to improve the characterization of the parts used as input substrate for device evolution. By using the kit, researchers will be able to report promoter and RBS activity in standard units (Standard Promoter Units, SPUs and Standard RBS Units, SRUs) enabling the growth of a library of characterized parts to draw on for assembling device variants. Additionally, I describe future work to specify a promoter standard ("fauxmoters") that will improve the functional composability of promoter parts. In Chapter 3, I describe a new microfluidic device, the Sortostat, that integrates a cell sorting chamber with a previously published microscope-mounted microfluidic chemostat. Researchers can use the Sortostat to apply morphological, time-varying, or other complex screens that might be associated with screening a system library. The bottleneck targeted by each chapter is marked in Figure 1-1.

Chapter 2. pSB1A10, A PoPS Input/Output Characterization and Screening Plasmid.

[The work presented in this chapter was carried out in collaboration with Josh Michener, Kelly Chang, Andrzej Wojcieszynski, and Felix Moser]

2.1 Summary

Directed evolution has proven to be a powerful tool in the biological engineer's arsenal, supporting the engineering of enzymes, fluorescent proteins, and many other parts. Applying directed evolution to more complicated devices and systems will require new tools for screening and selection. I designed and constructed a characterization and screening plasmid (pSB1A10) that could be used to characterize the input / output function of BioBrick parts and devices that use a transcription signal (RNA polymerase per second, PoPS) as input or output. I explored the functionality of pSB1A10 by characterizing a set of six BioBrick transcription terminators as well as a set of nine BioBrick inverters. I demonstrated use of pSB1A10 to screen an inverter library to isolate a functional inverter mutant. I attempted to combine two inverters in series and characterize them in pSB1A10 to explore the challenge of device composition. Finally, I implemented a new version of pSB1A10 to try to account for some of the shortcomings in the original design. Although these first-generation characterization and screening plasmids could be further improved, there is a clear need for

new tools and approaches to tuning biological parts and devices via library generation and screening.

2.2 Introduction

Biological engineers have constructed a number of multi-part synthetic biological systems that conduct logical operations on input signals [20, 21], produce oscillatory output signals [22], store memory [23], or produce desired products [24, 25]. However, very few of these synthetic biological systems worked as originally designed. The typical process of constructing a functional system involves a period of tuning the system to find a functional variant [26]. This tuning process has long been applied with success to the engineering of individual biological parts by directed evolution. For instance, researchers developing improved enzymes [2], transcriptional promoters [3], and fluorescent proteins [4] have made use of directed evolution to find variants that met their desired performance specifications. The directed evolution process involves generating large libraries (1E5-1E9 members) of part variants by using methods such as mutagenic PCR or *de novo* DNA synthesis followed by high-throughput screening or selection to isolate desirable mutants. A similar process may be applied to the construction of multi-part, synthetic biological systems, however it will require advances in our ability to assemble libraries of systems made from standard components and our ability to screen for the more complicated functions associated with these systems (Chapter 1).

Researchers have begun to explore the application of directed evolution techniques to multi-part devices. For instance, Yokobayashi et al successfully applied screening [19] and selection [18] to a library of inverters and isolated variants with desired properties. However, these proof-of-principle examples were not designed to support re-usability across many different parts or devices. A researcher applying these previous tools to screen a new system would need to repeat much of the construction and design work on the screening platform itself. Furthermore, screening using these previous tools was accomplished by a plate-based assay that relied on a qualitative measure a single reporter of the output of device function. More quantitative screens could be conducted using a FACS based screening assay with reporters for both input and output transcription rates.

Here I describe pSB1A10, a PoPS input / output characterization and screening plasmid. PoPS, or RNA polymerases per second, is a common signal carrier for transcription-based parts and devices [27]. My overall goals in designing pSB1A10 were to enable users to insert a BioBrick part or device into the plasmid and characterize the input / output function of the part or device as well as to enable the screening of libraries of devices based on their input / output functions. My design for pSB1A10 (Figure 2-1) consists of four components (1) an inducible PoPS generator device, (2) an input PoPS measurement device, (3) an insertion site for the test part or device to be characterized, and (4) an output PoPS measurement device. I used the paraBAD inducible expression system [28] as the inducible PoPS generator device, green fluorescent protein (GFP) and red fluorescent protein (RFP) as

indicators of input and output PoPS rates, respectively, and a BioBrick cloning site [29] flanked by RNAseE cleavage sites as the site for part or device insertion.

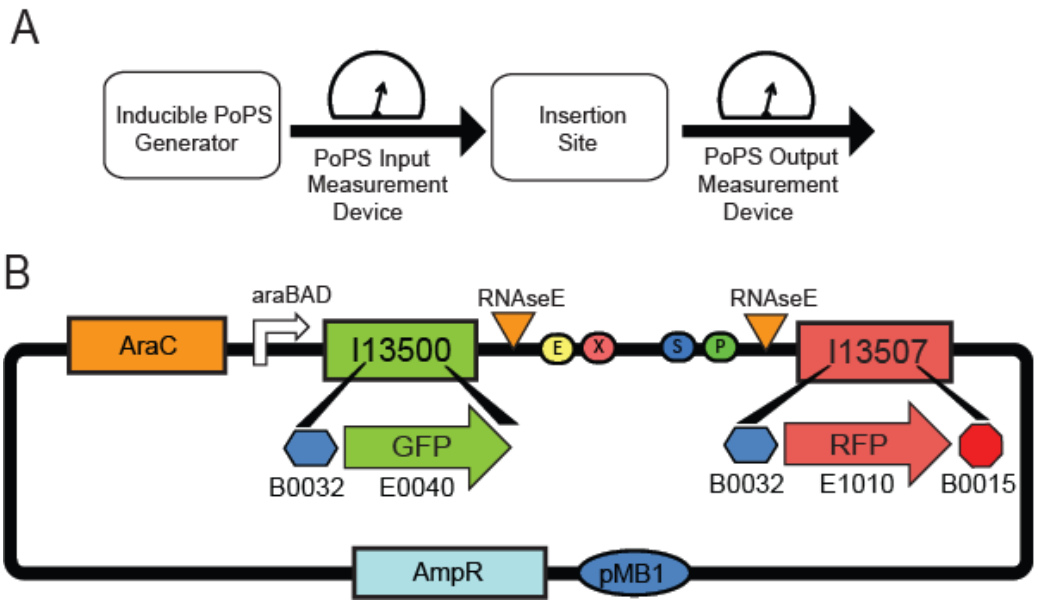


Figure 2-1 Schematic of the PoPS Input/ Output characterization and screening plasmid (pSB1A10)

(A) The design is made up of four components: (1) an inducible PoPS generator for providing a variable PoPS input into the test part or device, (2) a PoPS input measurement device for reporting the amount of PoPS input provided by the PoPS generator, (3) an insertion site for enabling cloning of a test part or device into pSB1A10, and (4) a PoPS output measurement device for reporting the amount of PoPS exiting the device. (B) The pSB1A10 construct contains: (1) an AraC expression cassette and paraBAD promoter that is inducible by arabinose as an inducible PoPS generator, (2) BBa_I13500 as the PoPS input measurement device, (3) The standard BioBrick cloning site flanked by RNAseE sites; the sites are used to insulate the test part or device mRNA from the reporter proteins mRNAs, and, (4) BBa_I13507 as the PoPS output measurement device.

In order to characterize a part or device using pSB1A10, users first insert the test part or device into the BioBricks cloning site on pSB1A10 via standard BioBricks assembly [30]. The assembled plasmid containing the test part or device is then transformed into the testing

strain (CW2553, Methods). The test part or device is characterized across a wide range of input PoPS levels by growing replicate cultures of the cells in media with different arabinose concentrations to induce the PoPS generator. The cells are grown to steady-state and the fluorescence intensity of GFP and RFP are measured using a flow cytometer to provide a measure of the input / output function for the test part or device.

I explored the functionality of pSB1A10 by attempting to characterize a set of six BioBrick transcription terminators as well as a set of nine BioBrick inverters (Figure 2). I "repaired" the functionality of a broken inverter by generating a library of inverters via mutagenic PCR and screening with pSB1A10 for a mutant with the appropriate function. I attempted to characterize the performance of two inverters in series using pSB1A10. Finally, based on my experiences using pSB1A10, I designed and tested a new version of a measurement and screening plasmid, pSB3K10.

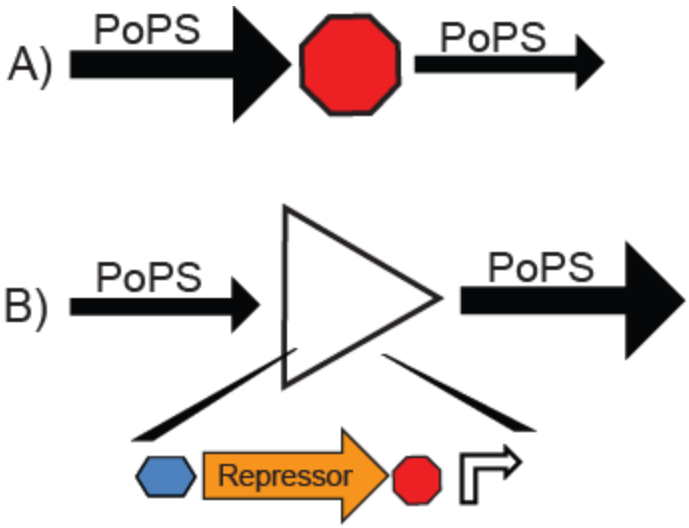


Figure 2-2 Examples of PoPS Input / Output devices

Inverters and transcription terminators are examples of devices and parts that send and receive PoPS as a signal. (A) A terminator receives an input PoPS signal and produces an output signal that is a fraction of the input. The fractional output depends on the efficiency of the terminator. A terminator that was 100% efficient would always output 0 PoPS, a terminator with 99% efficiency would output 1% of the PoPS input signal. (B) An inverter receives an input, performs a logical NOT operation, and produces a corresponding output. For example, the inverter here is shown converting a low PoPS signal into a high PoPS signal. An inverter is composed of four parts, a ribosome binding site (blue), the coding region for a repressor protein (orange), a transcription terminator (red), and the promoter regulated by the repressor (white).

2.3 Results and Discussion

2.3.1 pSB1A10 design & construction

I had five goals in designing pSB1A10. First, I wanted to make use of a common signal carrier for gene expression (such as polymerase per second, PoPS) [27] to enable any part or device that uses PoPS as an input or output signal to be characterized or screened without requiring the user of pSB1A10 to redesign the plasmid. Second, I wanted to include a reporter of the per cell input transcription rate (PoPS) into the test device to account for cell to cell variation of the activity of the arabinose-inducible paraBAD promoter in response to the inducer concentration. Third, I wanted to enable characterization and screening of parts and devices across a wide range of PoPS inputs and outputs. Fourth, I wanted reporters of input and output transcription rate that were insulated from the part or device being measured. In other words, the relationship between the measurable reporter output and the transcription rate (PoPS) should remain constant regardless of the device being tested, thus characterization should be comparable across different devices. Fifth, I wanted to conform to the BioBrick physical assembly standard for inserting parts or devices into pSB1A10 to allow compatibility with the large collection of existing parts and devices in the MIT Registry of Standard Biological Parts[31].

To meet my first design goal, I designed the PoPS output reporter device to convert a common signal carrier, PoPS, into a measurable fluorescence signal via expression of red fluorescent protein (RFP). I designed the PoPS input reporter device to measure PoPS

entering the test device by converting a PoPS signal into a measurable fluorescence signal via expression of green fluorescent protein (GFP). Because the input and output reporter devices were designed to detect PoPS, any device that accepts an input signal in PoPS and outputs a signal in PoPS can be characterized using pSB1A10 without needing to alter the plasmid. Examples of such parts and devices include transcription terminators and inverters (Figure 2-2).

To meet my second design goal, accounting for cell to cell variability in the activity of the input PoPS generator, I included an input reporter device (BBa_I13500) upstream of the insertion site that allows for a per cell measurement of the PoPS entering the test part or device based on measurement of the per cell concentration of GFP using a flow cytometer.

To meet my third design goal, I needed an inducible PoPS generator that provides a wide range of input PoPS. I chose the paraBAD promoter from the *E.coli* arabinose operon. This promoter has been shown previously to provide linear induction across nearly three orders of magnitude in a previously reported expression system [28]. To ensure linear induction of the paraBAD promoter Khlebnikov et al. removed the natively controlled arabinose transport gene (araE) and instead constitutively expressed araE from a plasmid. I chose the araE knockout strain (CW2553) containing a plasmid that constitutively expressed araE (pJat8) as the characterization strain [32].

To meet my fourth design goal, to insulate the reporter device performance from the specific part or device being tested, I included RNAseE cleavage sites flanking the insertion site [33]. The RNAseE cleavage sites were designed to insulate the stability of the GFP and

RFP mRNA from the part or device that is inserted between them. Insulation is necessary since inserted parts or devices may be transcribed as a polycistronic mRNA containing both the sequence of the part or device and the sequence of the reporter, thus potentially influencing the stability of the reporter mRNA. As an example, in response to a constant induction level the GFP (input) measurement could differ depending on the device inserted downstream of GFP, preventing reliable comparison of input measurements (Figure 2-3A). By including RNAseE sites I expect that the reporter mRNA will be cleaved from the test device mRNA forming separate secondary transcripts (Figure 2-3B) and increasing the insulation between reporter performance and the device being measured. If the lifetimes of the secondary transcripts are long compared to the lifetime of the original transcript then the system will behave as if the two mRNAs were produced independently. I conducted preliminary experiments to verify the efficacy of the RNAseE sites in pSB1A10 (Figure 2-4).

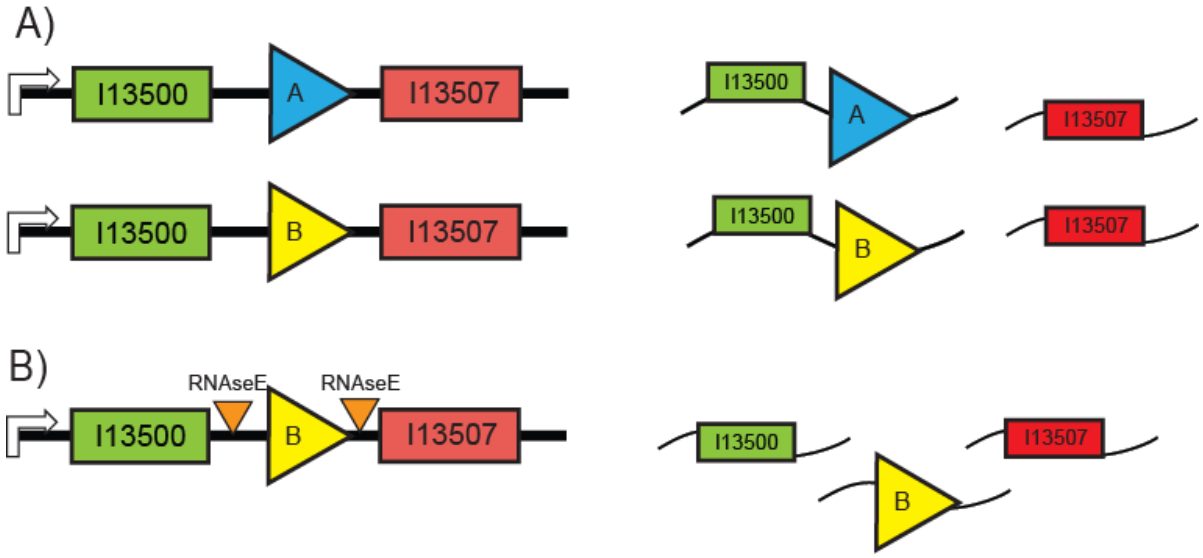


Figure 2-3 RNAseE sites were designed to insulate the reporter mRNA from the test part or device mRNA.

In order to make comparable measurements of the PoPS input and output function of a collection of test parts or devices, it is necessary that the PoPS input and output reporter devices function equivalently independent of the test part or device being measured. (A) In many cases (for instance with an inverter as shown here) one of the reporter genes is transcribed on the same mRNA as a portion of the test part or device. As a result of this polycistronic message, the PoPS input measurement device will produce mRNAs with different stabilities for different test devices (for instance, inverter A might stabilize the GFP mRNA, while inverter B might de-stabilize the mRNA). (B) In order to insulate the stability of the PoPS input reporter mRNA from the test part or device I included RNAseE sites so that the polycistronic mRNA would be cleaved by RNAseE. The cleavage event will produce secondary transcripts that are independent of the test part or device.

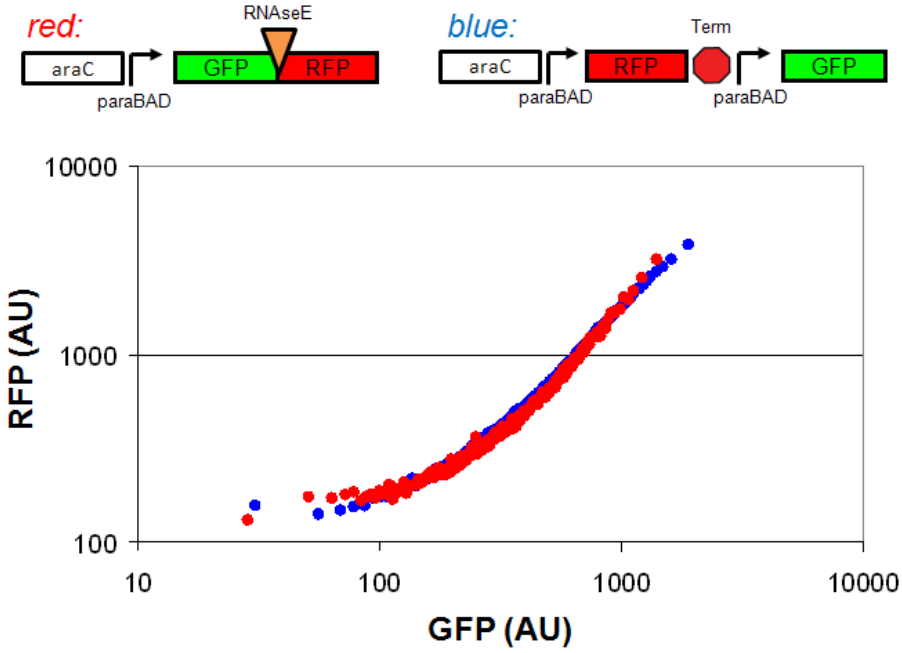


Figure 2-4 RNAseE sites help to insulate reporter mRNA stability

To test the efficacy of the RNAseE sites, I measured the GFP and RFP fluorescence (in arbitrary units, AU) from two constructs across six arabinose induction levels (Methods), *red:* GFP and RFP separated by an RNAseE site and expressed by the paraBAD promoter on a polycistronic mRNA and *blue:* GFP and RFP on separate mRNA transcripts each expressed from a separate paraBAD promoter. The fluorescence was measured on a flow cytometer and data from measuring six separate cultures at different arabinose induction levels was combined into a single data set. The plot shows the average GFP and RFP fluorescence for each of 100 logarithmically-spaced bins based on GFP fluorescence. The GFP and RFP fluorescence measurements are equivalent for the two constructs within the error of the measurements, suggesting that the RNAseE sites may help to insulate the reporter mRNA transcripts. The conclusion is further supported by previous results (not shown) comparing RFP on its own transcript to RFP on a polycistonic mRNA downstream of GFP with no RNAseE site between the coding regions. There I found that the polycistonic mRNA increased the expression of RFP, however the expression was not from the paraBAD promoter and therefore is not directly comparable to the results shown here. Further experiments to quantify the amount of cleaved and uncleaved transcript directly are needed to confirm the efficacy of the RNAseE sites.

To meet my fifth design goal, to enable compatibility with BioBrick parts, I chose to use the standard BioBrick cloning site for the insertion site in pSB1A10. Any part or device from the Registry of Standard Biological Parts can be easily inserted into pSB1A10 in a single cloning step [29].

2.3.2 Characterization of six BioBrick Transcription Terminators

I used pSB1A10 to measure the termination efficiency of a set of three BioBrick transcription terminators in the forward (BB_B0011, BBa_B0014, BBa_B0015) and reverse direction (BBa_B0021, BBa_B0024, and BBa_B0025). These terminators were chosen since they are used across many engineered biological systems built using BioBrick parts, in particular BBa_B0015 is one of the most widely used parts in the Registry (Randy Rettberg, personal communication). The termination efficiency was determined by measuring the per cell GFP and RFP fluorescence with a flow cytometer (Methods) after growing cells under a single induction level (0.003% arabinose). Terminator efficiencies were calculated by the following formula described previously [34]:

$$Termination\ Efficiency\ (per\ cell) = 1 - \frac{\frac{RFP_{term}}{GFP_{term}}}{\left(\frac{RFP_{control}}{GFP_{control}}\right)^{mean}} \quad \text{(Eq. 2.1)}$$

Where the subscript *term* refers to the measured GFP and RFP fluorescence when pSB1A10 contains a terminator and the subscript *control* refers to the measured fluorescence from pSB1A10 when no part is inserted in the insertion site (equivalent to 0% termination efficiency). The ratio of RFP to GFP fluorescence after background subtraction was

averaged across all cells in the population for the control (denoted by the *mean* superscript). The ratio of RFP to GFP fluorescence after background subtraction was calculated for each cell in the population of cells containing pSB1A10 with a terminator inserted. The mean measured termination efficiency across all cells in the population for each terminator is provided in table 2-1. Since I measured the termination efficiency for each individual cell I also determined the distribution of termination efficiencies across cells in the population (Figure 2-5).

Terminator BioBrick#	Terminator Efficiency
BBa_B0011	0.95
BBa_B0014	0.96
BBa_B0015	0.97
BBa_B0021	0.86
BBa_B0024	0.86
BBa_B0025	0.62

Table 2-1 Mean termination efficiency of six BioBrick terminators characterized with pSB1A10.

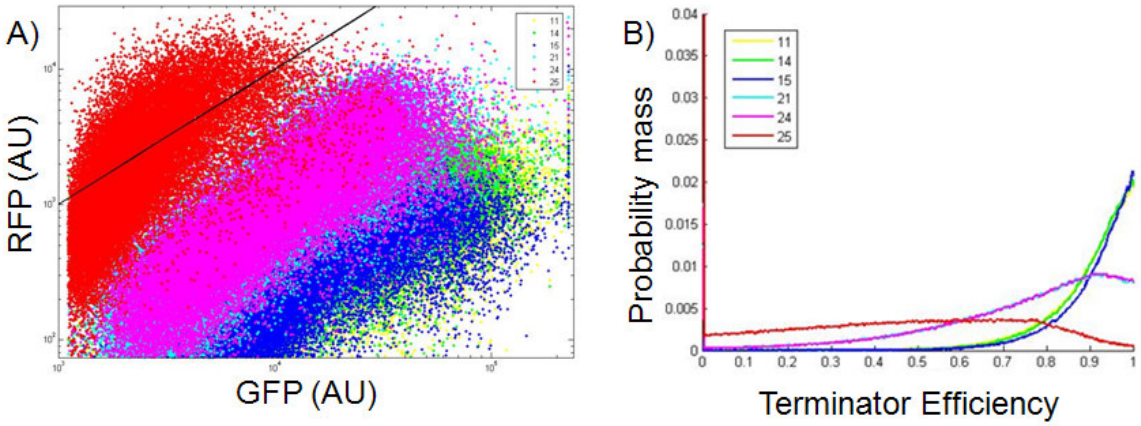


Figure 2-5 Characterization of the termination efficiency of six BioBrick terminators

(A) Dot plot of the flow cytometer measurements of GFP and RFP fluorescence from each cell in the population. The black line represents the best fit of the control pSB1A10 without a terminator inserted (in other words, termination efficiency = 0%). Terminators are expected to fall below the black line since for any given PoPS input (GFP) they are expected to produce a lower PoPS output (RFP) then the control with 0% termination efficiency. (B) The probability mass function (pmf) was calculated for the efficiency of each terminator.

In order to characterize the input / output relationship of a part or device using pSB1A10 I assume that the measured GFP and RFP fluorescence are directly related to the PoPS input and output of the part or device. However, in the case of transcription terminators this assumption is not fully justified. When a read-through event occurs (necessary for RFP to be expressed), a new hairpin is introduced to the 5' end of the RFP mRNA transcript due to the terminator hairpin itself being part of the transcribed mRNA. It is known that 5' hairpins are capable protecting transcripts from RNAseE cleavage [35], and as a result it is possible that the RFP mRNA transcript is not being cleaved at the engineered RNAseE site within pSB1A10. Without this cleaving event the insulation between the

particular part being measured and the downstream reporter (RFP) may be lost. This can have an unpredictable impact on the stability of the RFP mRNA transcript and will increase the error in terminator efficiency measurements. Future work is needed to better predict mRNA degradation rates from primary sequence or to design flanking sequences that might fix the degradation rates of mRNA at one predictable value [33].

2.3.3 Characterization of nine BioBrick Inverters

I measured the input / output transfer curves of a set of nine BioBrick inverters (Table 2). The nine inverters contained four different repressor proteins taken from natural biological systems: Tet repressor (BBa_C0040) from transposonTn10, c2 repressor (BBa_C0053) and Mnt (BBa_C0072) repressor from bacteriophage P22, and PenI (BBa_C0074) repressor from *Bacillus lichenformis*. Additionally each of the repressor proteins have a C-terminus degradation tag (LVA) to target the protein for degradation by ClpXP or ClpAP [36]. The tags were included to reduce the half-life of the repressor proteins so that protein levels in the cell will respond more rapidly to changes in transcription input to the inverter. Faster response times are often needed for the proper function of dynamic systems such as an oscillator [22].

I did not design or construct any of the inverters characterized here; all devices were received from the Registry of Standard Biological Parts [31]. For each type of inverter (Tet, c2, Mnt, or PenI-based) multiple variants were present in the Registry containing RBSs with different activities upstream of the repressor coding region (Table 2-2). I inserted each of the

nine inverters into pSB1A10 and measured the steady-state input and output PoPS level for each inverter in response to six induction levels (0%, 1E-6%, 3E-6%, 1E-5%, 3E-5%, 1E-4% arabinose) (Methods). Of the nine inverters tested I found that only one (BBa_Q04740) changed its output PoPS level across the range of induction levels tested (Figure 2-6 & 2-7). The remaining eight inverters were found to be unresponsive to different induction levels, remaining locked in either the low output (repressed) state or the high output (un-repressed) state (Table 2-2).

	RBSs upstream of the repressor coding region (relative RBS activity)			
	BBa_B0031 (0.01)	BBa_B0033 (0.07)	BBa_B0032 (0.38)	BBa_B0034 (1.0)
TetR	High Output BBa_Q01400	High Output BBa_Q03400		Low Output BBa_Q04400
P22 cII	Low Output BBa_Q01530			Low Output BBa_Q04530
PenI		High Output BBa_Q03740	High Output BBa_Q02740	Functional BBa_Q04740
Mnt				High Output BBa_Q04720

Table 2-2 Results of characterization of nine BioBrick inverters.

'High Output' indicates that independent of the input PoPS rate into the inverter the output of the inverter was always in the un-repressed state. 'Low output' indicates that independent of the input PoPS rate into the inverter the output of the inverter was always in the repressed state. 'Functional' indicates that across the PoPS input range provided by pSB1A10 the inverter transitioned from being un-repressed to repressed. Only one of the nine inverters was functional across the input range tested (BBa_Q04740).

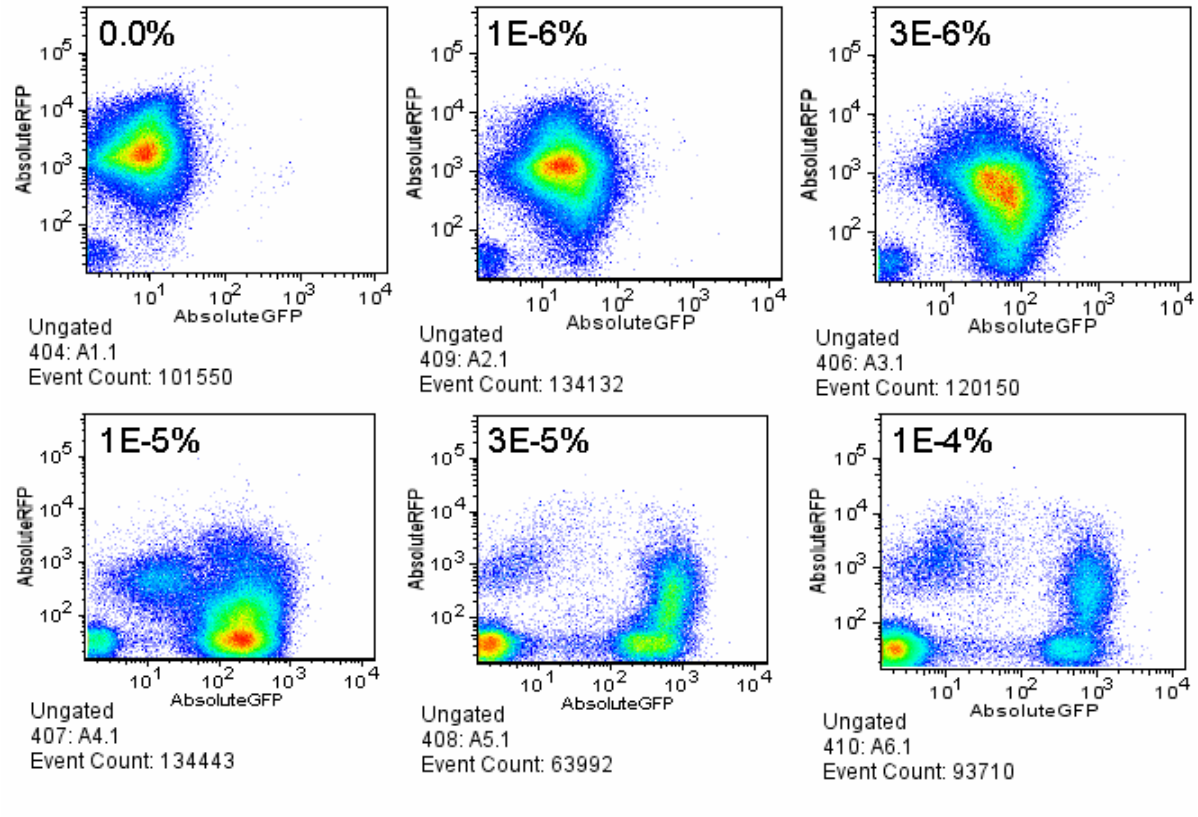


Figure 2-6 Characterization of BBa_Q04740

Dot plots of the GFP (x-axis, input) and RFP (y-axis, output) fluorescence measurements of the inverter BBa_Q04740 characterized in pSB1A10 across six different arabinose induction levels (arabinose percent w/w is shown in figure).

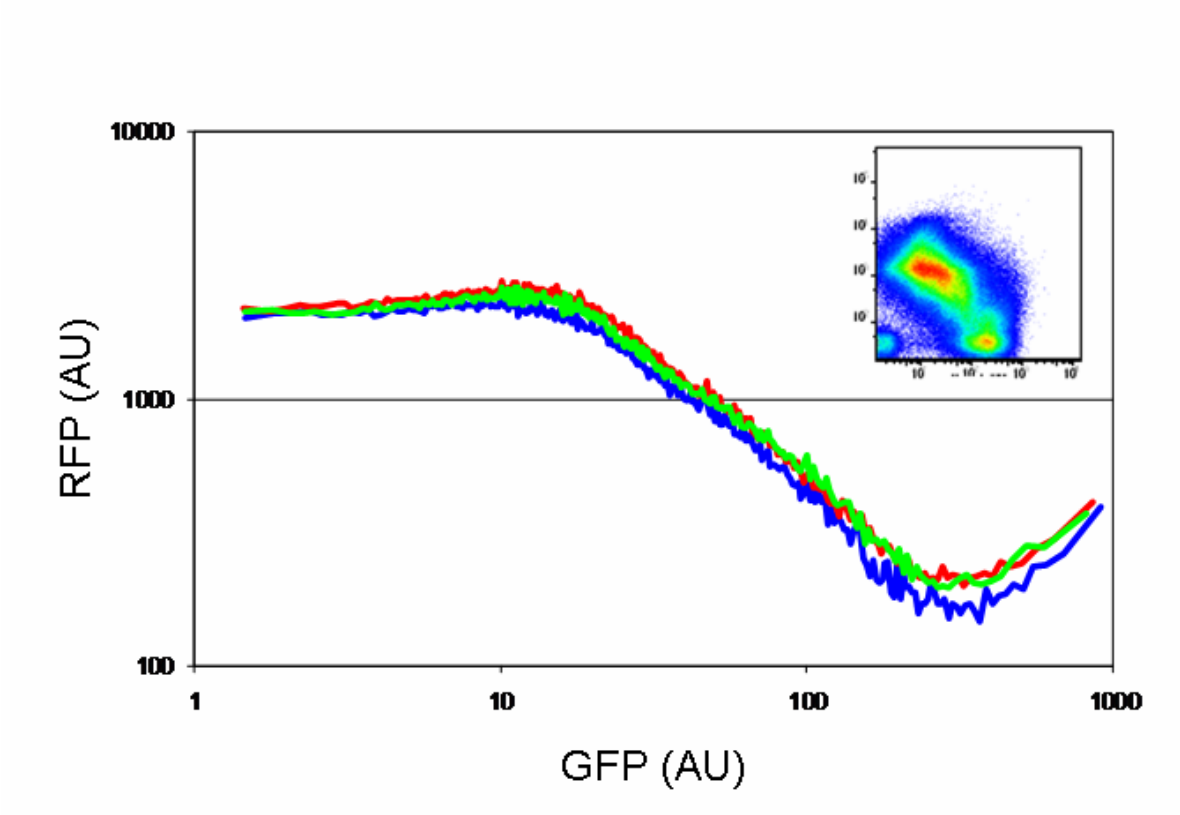


Figure 2-7 Summary of triplicate experiments characterizing the inverter BBa_Q04740.

The dot plot in the upper right is the concatenation of the six dot plots shown in figure 2-6. This concatenated data was then grouped into 100 logarithmically-spaced bins based on GFP fluorescence and the mean GFP and RFP fluorescence was calculated for each bin to produce the simplified transfer curves shown in the main graph. Each line on the transfer curve represents a replicate characterization of BBa_Q04740.

It is not surprising that eight of nine inverters did not properly function in the context of pSB1A10. In their native systems, the repressor proteins used in the inverters are present at different copy numbers and in different genetic backgrounds and culture conditions. We should expect that many parts and devices harvested from natural systems will need to be

"tuned" in order to perform appropriately in our engineered systems [26]. In order to demonstrate the use of pSB1A10 to tune inverters via library screening I repaired one of the eight non-functional BioBrick inverters: BBa_Q04400, the Tet repressor-based inverter.

2.3.4 Repair and Reuse of a Tet repressor-based inverter

The details of the characterization of the failed inverter variants (Table 2-2) are informative in deciding the appropriate strategy for tuning inverters to function in pSB1A10. In particular the characterization results of the three variants of the Tet repressor-based inverter (BBa_Q01400, BBa_Q03400, and BBa_Q04400) show that by changing the activity of the RBS in the inverter the function of the device changes from a state where it is un-repressed across all induction levels (BBa_Q01400, BBa_Q03400) to a state where it repressed across all induction levels (BBa_Q04400). Furthermore, these results make intuitive sense in light of the measured activities of the RBSs in the three Tet repressor-based inverters. An RBS with a higher activity will produce more repressor protein across all induction levels. Thus, I expect a more active RBS such as the one in BBa_Q04400 to lead to inverters locked in the repressed state due to over-expression of repressor protein, while weaker RBSs such as those in BBa_Q01400 and BBa_Q03400 to be more likely to lead to inverters locked in the un-repressed state because of insufficient expression of repressor proteins. The results of the characterization of the three Tet repressor-based inverters suggested that both the Tet repressor (BBa_C0040) and the Tet repressor-regulated promoter (BBa_R0040) were functioning properly as the promoter could be both repressed

(BBa_Q04400) and un-repressed (BBa_Q01400 and BBa_Q03400). Thus, I expected that repair of the inverter could be accomplished by tuning the activity of the RBS.

The other broken inverters (Mnt and p22 cII repressor-based) showed either a low output with a weak RBS (BBa_Q01530) or high output with a strong RBS (BBa_Q04720). Thus it is possible that these inverters have different flaws than an incorrect expression level of the repressor protein, such as a dysfunctional repressor protein (BBa_Q04720) or a dysfunctional promoter (BBa_Q01530). Depending on the cause of the device failure these problems might be harder to solve via library generation and screening than tuning expression level. I attempted to repair all three inverters (not shown), however I was only successful in tuning the performance of the Tet repressor-based inverter, BBa_Q04400.

In order to repair BBa_Q04400 I generated a library of inverters by amplifying the BBa_Q04400 DNA with a mutagenic PCR reaction using primers that maintained the standard BioBrick restriction sites at the ends of the DNA sequence of each library member (Methods). I digested the inverter library with EcoRI and PstI and inserted the library into pSB1A10 in a single cloning reaction. The transformants were maintained in liquid culture rather than plated in order to maintain a large library size (~1E5 mutants). I then conducted two rounds of screening on the inverter library (Figure 2-8). In the first round I grew the cells with low induction (no arabinose) and screened for cells that had a low input PoPS (low GFP signal) and a high output PoPS (high RFP signal) using a FACS machine to sort the cells that met my criteria into a sub-library. I conducted a second round of screening on the sub-library of cells generated in the first round; I grew the cells in the presence of high

induction (1E-4% arabinose) and screened for cells that had a high input PoPS (high GFP signal) and a low output PoPS (low RFP signal) again using a FACS machine to sort promising candidates into a separate sub-library.

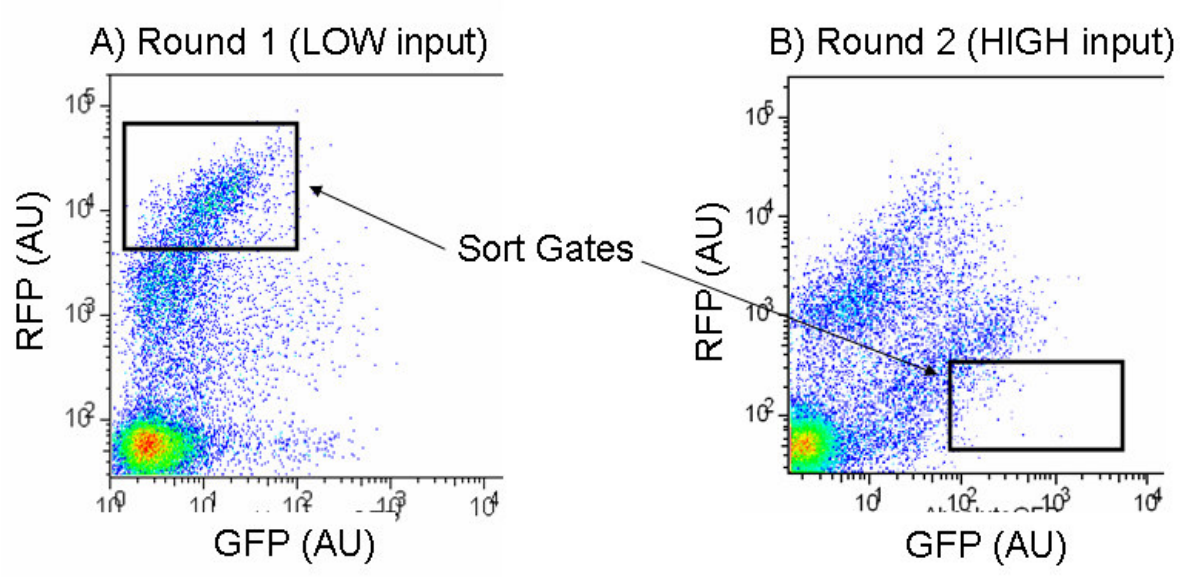


Figure 2-8 Two rounds of screening on a FACS machine of an inverter library generated via mutatgenic PCR of the non-functional inverter, BBa_Q04400.

In the first round of screening (A) the cells are gown in media with 0% arabinose (low input) and cells are sorted using a FACS machine based on the sort gate shown. The sort gate defines a region of cells with a low input (1X-10X background GFP) and a high output (100X – 1000X background RFP). In the second round of screening (B) the subpopulation of cells screened in round 1 are grown in media with 10E-4% arabinose (high input) and are selected based on a sort gate for high input (100X-1000X background GFP) and low output (1X-10X background RFP).

The final sub-library after two rounds of screening was plated to isolate individual mutants as single colonies, and five of these mutants were tested across six different arabinose induction levels (0%, 1E-6%, 3E-6%, 1E-5%, 3E-5%, 1E-4% arabinose). One

mutant was found that had a characteristic input / output transfer curve for an inverter (Figure 2-9), this inverter was given a new BioBrick part number, BBa_Q04401.

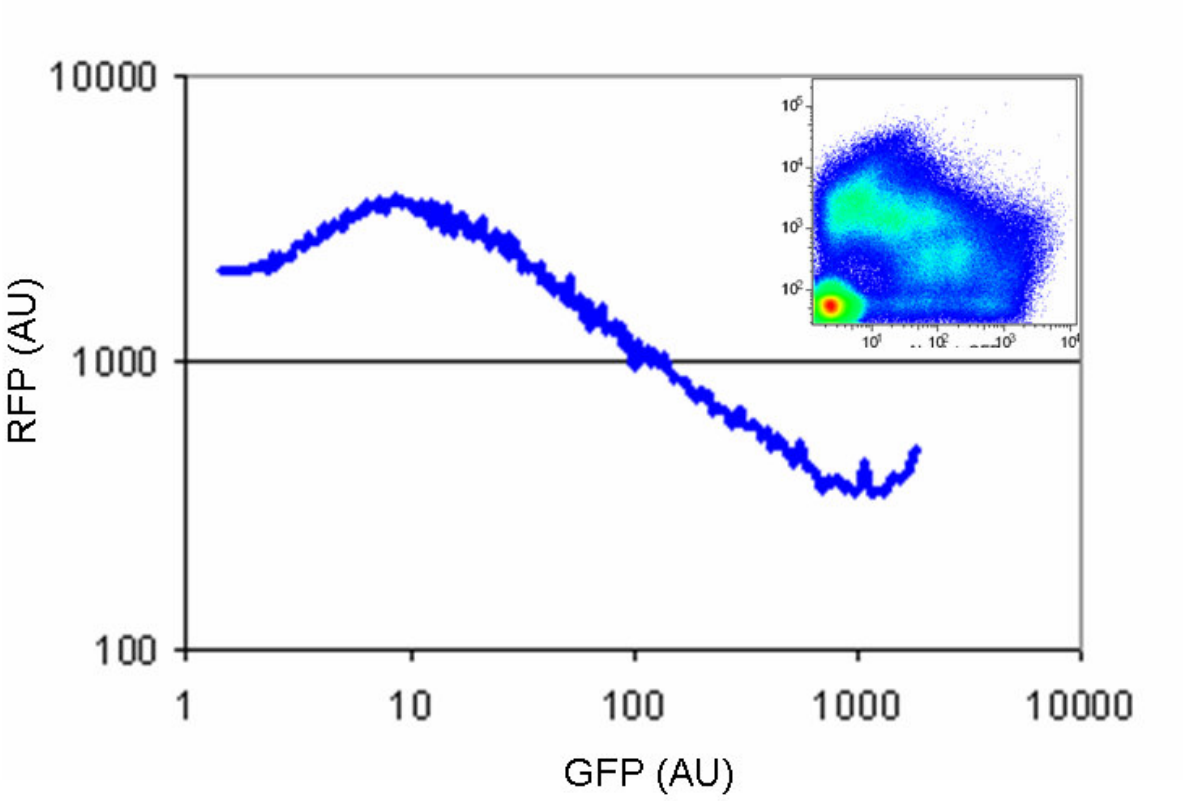


Figure 2-9 Characterization of BBa_Q04401, a successfully repaired inverter.

The dot plot in the upper right is the summary data of the GFP (input) and RFP (output) fluorescence measurements of the inverter BBa_Q04401 characterized in pSB1A10 across six different arabinose induction levels (0%, 1E-6%, 3E-6%, 1E-5%, 3E-5%, 1E-4% arabinose). This summary data was then grouped into 100 logarithmically-spaced bins based on GFP fluorescence and the mean GFP and RFP fluorescence was calculated for each bin to produce the simplified transfer curve shown in the main graph.

I sequenced BBa_Q04401 and found a single base pair mutation in the RBS upstream of the repressor coding region (the mutant RBS was given a new BioBrick part number,

BBa_B0064). I tested the activity of this RBS and found it to be 36% as active as the original RBS in BBa_Q04400 and about 5X as active as the weaker RBS in BBa_Q04340 (not shown). These results confirmed my intuition that tuning of the expression level could repair this inverter. The process also suggested that I may be able to produce libraries containing more functional mutants by combining collections of functional elements (for instance, inserting a collection of RBSs of varying strength into the inverter) rather than by generating libraries via random mutagenesis of DNA. However, previous work by Yokobayashi et al to tune a cI-based inverter by generating a library at the sequence level yielded several functional variants with mutations that inserted stop codons within the cI gene resulting in truncated repressor proteins [19]. These types of mutants would be less likely to be made in a parts-based approach to generating part libraries, thus it remains an open question whether generating libraries via mutagenesis at the sequence level or via combining collections of functional elements would yield libraries with more functional variants.

BBa_Q04401 was successfully re-used by the MIT 2006 iGEM team [37] to invert the signal of a stationary phase promoter (Figure 10). This promoter was present on pSB1A2 which has the same origin of replication and thus a similar copy number to pSB1A10. In general, re-using devices in a similar context to the context in which they were tuned will increase the likelihood of successful device function.

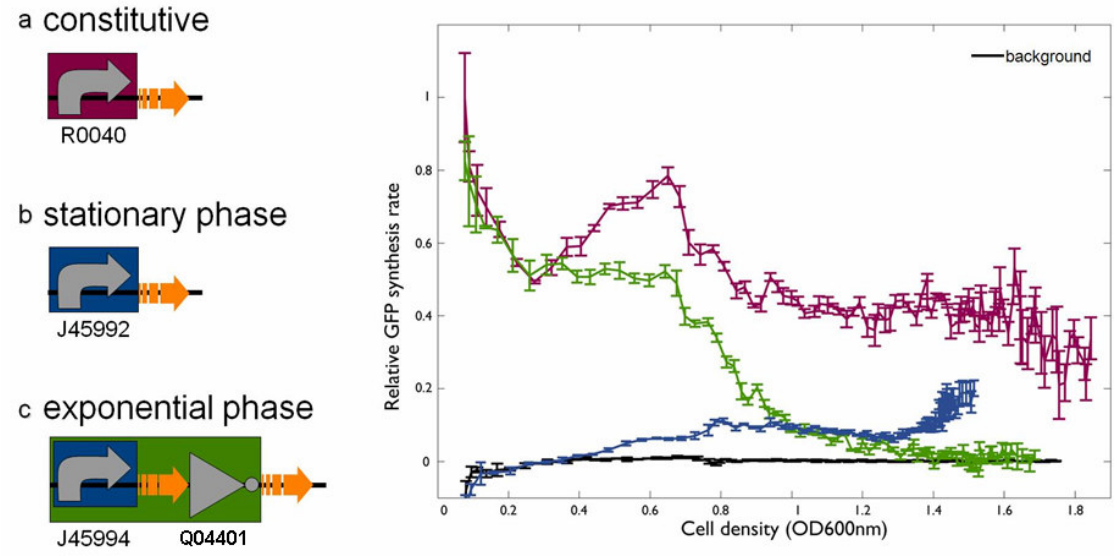


Figure 2-10 Successful reuse of inverter BBa_Q04401 by the MIT iGEM team in an engineered system for controlling bacterial odor.

The blue line represents the promoter activity measured by GFP synthesis rate per cell in a growing culture from a promoter (BBa_J45992) that increases in activity in stationary phase (notice the increase in activity from OD 0.4 to 0.8). The green line represents the same promoter upstream of inverter BBa_Q04401. The "inverted" stationary phase promoter now shows a high activity in exponential phase (OD < 0.6) and upon entry into stationary phase the activity drops significantly (down to background levels by OD = 1.4). Figure courtesy of Reshma Shetty [13].

2.3.5 Testing inverters in series, BBa_Q04401 and BBa_Q04740

After successfully screening for a functional Tet repressor-based inverter, I now had two functioning inverters (BBa_Q04401 and BBa_Q04740). A central challenge in system construction is the composition of individual devices into a larger system that works as predicted. For instance, would the two functioning inverters work properly when placed in series? Though other researchers have previously combined inverters in series (Table 3), predicting the function of the multi-inverter system *a priori* remains challenging. For instance, based on the transfer curves for each inverter (Figure 2-7 and 2-9) I could make a rough estimate of whether the inverters would function properly in series (Methods). I found that the inverters should work in one order (BBa_Q04401 preceding BBa_Q04740), but not in the reverse order.

I assembled the inverters in both orders and attempted to insert the pair into pSB1A10. I found that when Q04401 preceded Q04740 the output of the combined device was always a low signal independent of the input PoPS, thus my prediction based on the individual transfer curves was inaccurate. Additionally, the cells grew significantly slower (growth rate reduced by ~25%) than cells containing only a single inverter. I expect this difference in growth rate affects the steady-state concentrations of repressor proteins, leading to system functioning that is harder to predict. In the reverse order (Q04740 preceding

Q04401) I was unable to successfully clone the construct after many attempts. I concluded that pSB1A10 containing the inducible PoPS-generator device, two fluorescent protein PoPS reporter devices, as well as two inverters in series placed too much of a load on cellular resources, particularly on a high copy plasmid. To address this problem, I designed a new version of pSB1A10.

Promoter / Repressor 1	Promoter / Repressor 2	Direction of coupling	Replication Origin	Antibiotic resistance	Strain	Ref
Ptrc-2 / lacI	P_Ls1con / temp-sensitive λ repressor (*cIts*)	Both	ColE1 (pBR322 vector)	AmpR	JM2.300	[23]
Ptrc-2 / lacI	$P_L$tetO-1 / Tet repressor (*tetR*)	Both	ColE1 (pBR322 vector)	AmpR	JM2.300	[23]
$P_L$tetO-1 / Tet repressor (*tetR*)	Plac / lacI	1 → 2	p15A	KanR	Transfor Max EPI300	[38]
Plac / lacI	$\lambda_{P(R-O12)}$ / *cI*	1 → 2	p15A	KanR	Transfor Max EPI300	[38]
$P_L$lacO1 / lacI-lite	$P_L$tetO1 / tetR-lite	1 → 2	pSC101	AmpR	MC4100	[22]
$P_L$tetO1 / tetR-lite	λP_R / cI-lite	1 → 2	pSC101	AmpR	MC4100	[22]
λP_R / cI-lite	$P_L$lacO1 / lacI-lite	1 → 2	pSC101	AmpR	MC4100	[22]
Plac(on colE1) / lac (on chromosome)	Ptet / tetR-eCFP fusion (on colE1)	1 → 2	Chromosome / colE1	AmpR	JM101	[39]
Plac / lacI (on p15A)	$\lambda_{P(R-O12)}$ (on ColE1) / *cI* (on P15A)	1 → 2	p15A / ColE1	KanR(p15A) / AmpR (ColE1)	DH5α	[19]

Table 2-3 Listing of inverters previously coupled in synthetic biological systems.

2.3.6 pSB3K10, a low copy, simplified PoPS input / output characterization and screening plasmid

I re-designed the characterization and screening plasmid by (1) moving to a low copy plasmid in order to reduce the load on the cellular resources, (2) removing the input measurement device to simplify the measurement of device performance and to further reduce the load on cellular resources, and (3) changing to an inducible PoPS generator with less strain and plasmid requirements. The new design uses BBa_F2621 as an inducible PoPS generator, GFP as the reporter of PoPS output, and a p15A origin of replication that results in approximately a 10X lower copy number than pSB1A10. I tested pSB3K10 (Figure 10) by inserting the BBa_Q04401 inverter that I had previously tuned to function in pSB1A10, however I found that the inverter did not function properly in pSB3K10 and was locked in the low output state across a range of induction levels (data not shown). It is likely that I would need to re-tune the inverter to function in this new context (new PoPS input range, new copy number of the plasmid). To support explore whether inverter libraries with more functional variants might be made by inserting libraries of parts, I have designed a new construct for tuning inverters by inserting part libraries, rather than making mutations at the sequence level (Chapter 5).

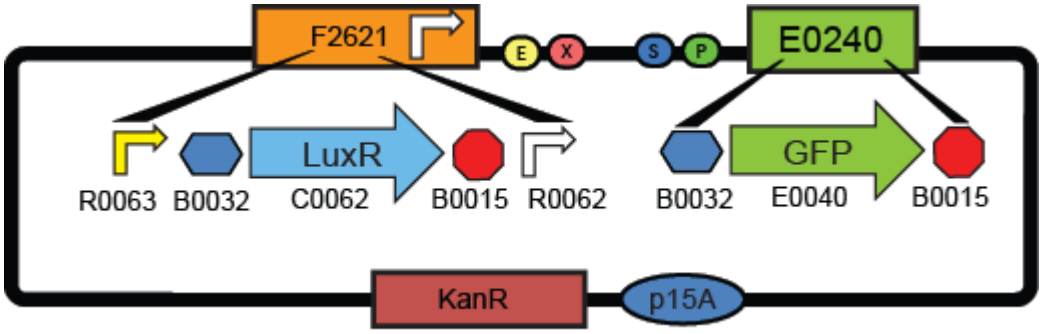


Figure 2-11 Schematic of pSB3K10.

A simplified version of the original measurement and screening plasmid was designed. This design does not require any particular strain background, and also will place a lower demand on cellular resources than pSB1A10 due to the lower copy number of the plasmid and having only one PoPS reporter device (BBa_E0240).

2.4 Conclusions

I explored the functionality of the characterization and screening plasmid, pSB1A10, by characterizing a set of six BioBrick transcription terminators as well as a set of nine BioBrick inverters. I found all but one of the inverters did not function over the range of PoPS inputs provided by the inducible PoPS generator device in pSB1A10. The dysfunction of eight of the inverters is not surprising; in their native systems the repressor proteins are present in different copy numbers, genetic backgrounds, and culture conditions. We should expect that many parts and devices harvested from natural systems will need to be "tuned" in order to perform appropriately in our engineered systems [26].

In order to demonstrate the use of pSB1A10 to tune inverters via library screening I successfully repaired one of the eight non-functional BioBrick inverters: BBa_Q04400, the Tet repressor-based inverter. I repaired the functionality of BBa_Q04400 by generating a

library of inverters via mutagenic PCR and screening with pSB1A10 for a mutant with the appropriate function. A mutant was isolated, characterized, and reused by others in a newly engineered system. The approach of tuning parts in a format (such as the BioBrick standard) that allows for easy reuse by others will improve our ability to reliably construct multi-part synthetic biological systems.

I attempted to combine two inverters in series as well as to characterize BBa_Q04401 (the repaired inverter) in a new lower-copy characterization plasmid. Both these experiments failed and point to future work that is needed to improve the reliability of device tuning and composition. First, part and device definitions are needed that specify the characteristics of parts and devices needed to enable reliable composition of these components into multi-part synthetic systems [13]. Second, future work is needed to build tools for tuning parts or devices to meet these defined standards for interoperability, reliability, or other desired part characteristics. In Chapter 6 I outline further research in support of this second goal.

Chapter 3. Measurement kits and reference standards for BioBrick promoters and ribosome binding sites.

[This chapter is based on a manuscript that I co-wrote with Drew Endy (Kelly et al., submitted). The experimental work was done in conjunction with Adam Rubin. Experiments were run at independent laboratories by Caroline M Ajo-Franklin, John Cumbers, Michael J. Czar, Kim de Mora, Aaron L Glieberman, and Dileep D Monie]

3.1 Summary

3.1.1 Background

The engineering of many-component, synthetic biological systems would be made easier by the development of collections of well-characterized, standard biological parts. The first examples of standard biological parts are BioBrick parts. These parts adhere to a technical standard that facilitates reliable physical assembly of parts into multi-component engineered biological systems, however the behavior of these systems remains difficult to predict *a priori*. Researchers could better predict the behavior of systems composed of standard biological parts if the individual parts were well characterized. Standard tools, techniques, and units of measure will be needed to support the characterization of parts by independent researchers across many laboratories.

3.1.2 Results

We designed, built, and tested measurement kits for characterizing the *in vivo* activity of BioBrick promoters and ribosome binding sites (RBS). The kits contain measurement tools for determining promoter and RBS activity based on expression of green fluorescent protein. The kits also include a specific promoter and RBS selected as reference standards. By reporting the activity of a user-specified promoter or RBS relative to the activity of the reference promoter or RBS, independent researchers can calibrate their measurements and report them in newly defined, comparable units: standard promoter units (SPUs) and standard RBS units (SRUs). We demonstrated the utility of the kits by measuring the activities of a collection of promoters and RBSs across a range of conditions and procedures. Finally, we distributed four promoters to six independent laboratories for characterization using the promoter measurement kit. The low level of variability in measured promoter activity (coefficient of variation ~ 0.113) across laboratories demonstrates that the kits provide a first mechanism for independent laboratories to report comparable measurements of the functioning of standard biological parts.

3.1.3 Conclusions

The measurement kits will be distributed within the annual BioBrick parts distribution and we encourage researchers to use the kits for characterization of BioBrick promoters and RBSs. Users of the kits can share data describing the activity of promoters and RBSs via the Registry of Standard Biological Parts

(http://partsregistry.org/measurement). Researchers working to improve measurement kits or reference standards can participate in open technical discussion via the BioBricks Foundation (http://biobricks.org/standards).

3.2 Introduction

The engineering of many-component, synthetic biological systems would be made easier by developing collections of well-defined, standard biological parts [10, 21, 40-43]. For example, standard components have been instrumental in managing complexity in most mature engineering fields by allowing engineers to reliably predict the function of large numbers of interacting components [44]. However, it is an open question whether the overwhelming complexity of living systems will prevent biological engineers from achieving similar design capabilities. To help explore this question the MIT Registry of Standard Biological Parts maintains and distributes thousands of BioBrick standard biological parts [45]. BioBrick parts are the first example of standard biological parts that have been designed to enable reliable physical assembly of individual BioBrick parts into larger multi-component systems [40].

The ability of engineers to predict the behavior of engineered biological systems assembled from standard biological parts would be made easier if the functioning of these parts was well characterized [46-48]. Unfortunately, BioBrick parts are the only existing standard biological parts; most BioBrick parts remain to be quantitatively characterized. Part characteristics worth measuring vary by part type but typically include: static performance,

dynamic performance, genetic stability, or other characteristics [38, 49]. Characterization of parts might be carried out by dedicated, centralized facilities or by a distributed community of researchers. In either case, standard tools, techniques, and units of measure will be needed to reduce variability in the characterization of parts across independent researchers at many facilities. Reporting the activities of parts in standard units will allow future users of the parts to more confidently combine parts that have been characterized by an external group with parts that they themselves design and characterize.

Previous efforts to make standardized measurements of cellular functions highlight the challenges of taking consistent measurements across independent laboratories. For example, an analysis of 80 published papers in which researchers used beta-galactosidase (β-gal) activity as a measure of gene expression found at least six different protocols were used to measure enzymatic activity [50]. In addition, nearly all activities were reported in "Miller units" even though in several cases there were differences in the substrates used to quantify enzymatic activity (CPRG or ONPG), the experimental conditions (pH and temperature for the assay), and even the units of the Miller unit (nmol/min or μmol/min) [51]. Differences in conditions such as using either CPRG or ONPG as a substrate for enzymatic assays lead to incompatible results [52], thus Miller Units should generally not be considered comparable unless they have been calibrated against a common reference standard [50]. As a second example, comparing microarray data between independent studies is often impossible due to differences in array platforms, algorithms for image and data analysis, gene probe size, and many other parameters [53]. To address this challenge, the microarray community is

currently developing a set of external RNA reference standards to calibrate results across different array platforms [54, 55] and previously developed a community-wide specification, the Minimum Information About a Microarray Experiment (MIAME) definition, that can be used to standardize data input and reporting [56]. The biological engineering community should take a similar approach by developing and sharing standard methods, tools, and reference materials in support of the characterization of biological parts.

Reference standards have been used to account for variability in measurement procedures between independent workers in other engineering fields. For example, the British Association standard ohm resistor is a widely accepted reference standard [57] (Figure 3-1). Official copies of this reference standard for resistivity equivalent to one ohm were distributed to telegraph engineers, experimental physicists, and other researchers so that they could report their resistivity measurements relative to the resistivity of the reference standard (in other words, in standard units of ohms). The design of the standard ohm resistor was the end product of a long process of community discussion and debate that started with *ad hoc* reference standards shared among smaller groups of telegraph engineers [58]. For example, one of the earliest reference standards was used by telegraph engineers to demonstrate significant variability in the resistivity of 'pure' copper wire manufactured by different wire factories [58]. Prior to this *ad hoc* standard, differences in the resistivity of wires had been blamed on poor measurement tools or procedures rather than on the wire itself. The discovery of differences in the resistivity of supposedly pure copper wires

established demand for greater quality control in wire manufacturing, and in turn led to the production of improved wires with more consistent resistivity.

Figure 3-1 The British Standard Ohm resistor is an early example of a reference standard.

This resistor was designed in 1873 by the British Association for the Advancement of Science as the physical object representing the newly created official unit of resistivity, the ohm. Thousands of these "resistivity boxes" were built and shipped to telegraph engineers, physicists, and other researchers to ensure they could report their resistivity measurements relative to the resistivity of the standard resistor (in other words, in units of ohms). Instructions for use are shown on the lid of the box. Image courtesy of the Science & Society picture library [59].

Today, would-be biological engineers do not have sets of well characterized, off-the-shelf biological parts that can be reliably combined to produce larger systems with predictable performance [16]. The lack of well-characterized parts is at least partially due to

differences in the experimental conditions or measurement instruments used to characterize parts across different laboratories. As in the case of the standard ohm resistor, reporting measurements relative to reference standards for biological functions might help to account for some of the variability in measurement techniques across different laboratories and allow for more reliable characterization of parts. Better accounting for variability in characterization methods will also help to reveal sources of variability inherent in the behavior of the parts themselves, such as cross-talk between parts [60] or stochastic noise in gene expression [38, 39, 61].

Here, we describe first generation measurement kits for characterizing the activity of BioBrick promoters and ribosome binding sites (RBSs) (Figure 2). Our overarching objective for the kits was to enable independent researchers to make comparable measurements of promoter and RBS activity in standard units. We chose to develop kits for measuring promoter and RBS parts since these part types are ubiquitous in engineered biological systems, well-understood, and practically useful to biological engineers [62, 63]. To enable other researchers to make use of our kits we developed instructions and a parts list for both the promoter measurement kit (Figure 3-3 and Table 3-1) and the RBS measurement kit (Figure D-1 and Table D-1).

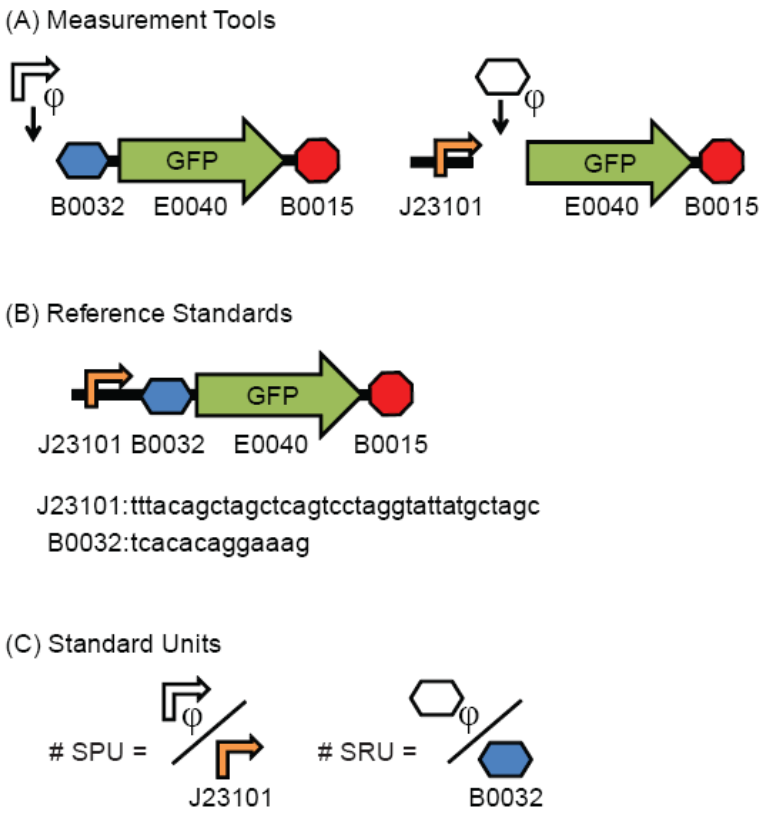


Figure 3-2 Overview of using the measurement kit

Standard measurement tools are used to characterize promoters and RBSs relative to reference standards, and results are reported in standard units. (A) Insertion of a user-specified promoter or RBS (both are shown in white) upstream of the GFP reporter devices to build the promoter and RBS test constructs. The promoter is inserted upstream of a fixed RBS (BBa_B0032), GFP coding region (BBa_E0040), and transcription terminator (BBa_B0015). The RBS is inserted downstream of a fixed promoter (BBa_J23101) and upstream of a GFP coding region (BBa_E0040) and a transcription terminator (BBa_B0015). **(B)** The reference standards are the promoter BBa_J23101 and RBS BBa_B0032. For simplicity we used the promoter reference standard (BBa_J23101) as the fixed promoter in the RBS test construct shown in (A), and we used the RBS reference standard (BBa_B0032) as the fixed RBS in the promoter test construct. As a result, the two reference standards can be contained in the same reference standard construct, reducing the number of components in the kit. **(C)** Kit users compare the activity of the promoter or RBS they are testing to the strength of the reference standard promoter (BBa_J23101) or RBS (BBa_B0032), respectively. This relative value allows measurements to be reported in common units of Standard Promoter Units (SPUs) and Standard RBS Units (SRUs).

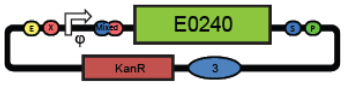
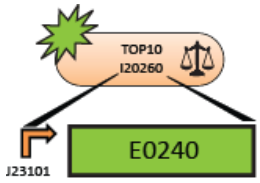
Notation	Name (BioBrick Part #)	Description / Function
	GFP reporter device (BBa_E0240)	The **GFP reporter device** contains an RBS, GFP coding region, and transcription terminator. It converts a transcription input into molecules of GFP.
	Backbone plasmid (pSB3K3-P1010)	The **backbone plasmid** contains an expression cassette for ccdb toxin that serves as a counter-selectable marker as well as the p15A origin of replication and a kanamycin resistance gene.
	Test promoter φ	The **test promoter φ** is the user-specified promoter to be measured.
	Promoter test construct	The **promoter test construct** contains the **test promoter** φ upstream of the **GFP reporter device** within the pSB3K3 **backbone plasmid.** The rate of GFP expressed from this construct is used to measure the activity of the test promoter.
	Reference standard construct (BBa_I20260)	The **reference standard construct** is identical to the **promoter test construct** except it contains the reference standard promoter (BBa_J23101). The rate of GFP expressed from this construct is used to measure the activity of the reference standard promoter.
	Preparative primers (BBa_G1000, BBa_G1001)	The **preparative primers** are used to amplify the pSB3K3 **backbone plasmid**.
	TOP10 (BBa_V1009)	*E. coli* TOP10 is used as the standard strain for kit measurement experiments.

Table 3-1 Components of promoter measurement kit

The nomenclature listed is used to describe the promoter measurement kit components within the text. The notation for each component serves as a key for the kit instructions (Box1). More detailed information about each component such as the DNA sequence can be found at the Registry of Standard Biological Parts (http://partsregistry.org).

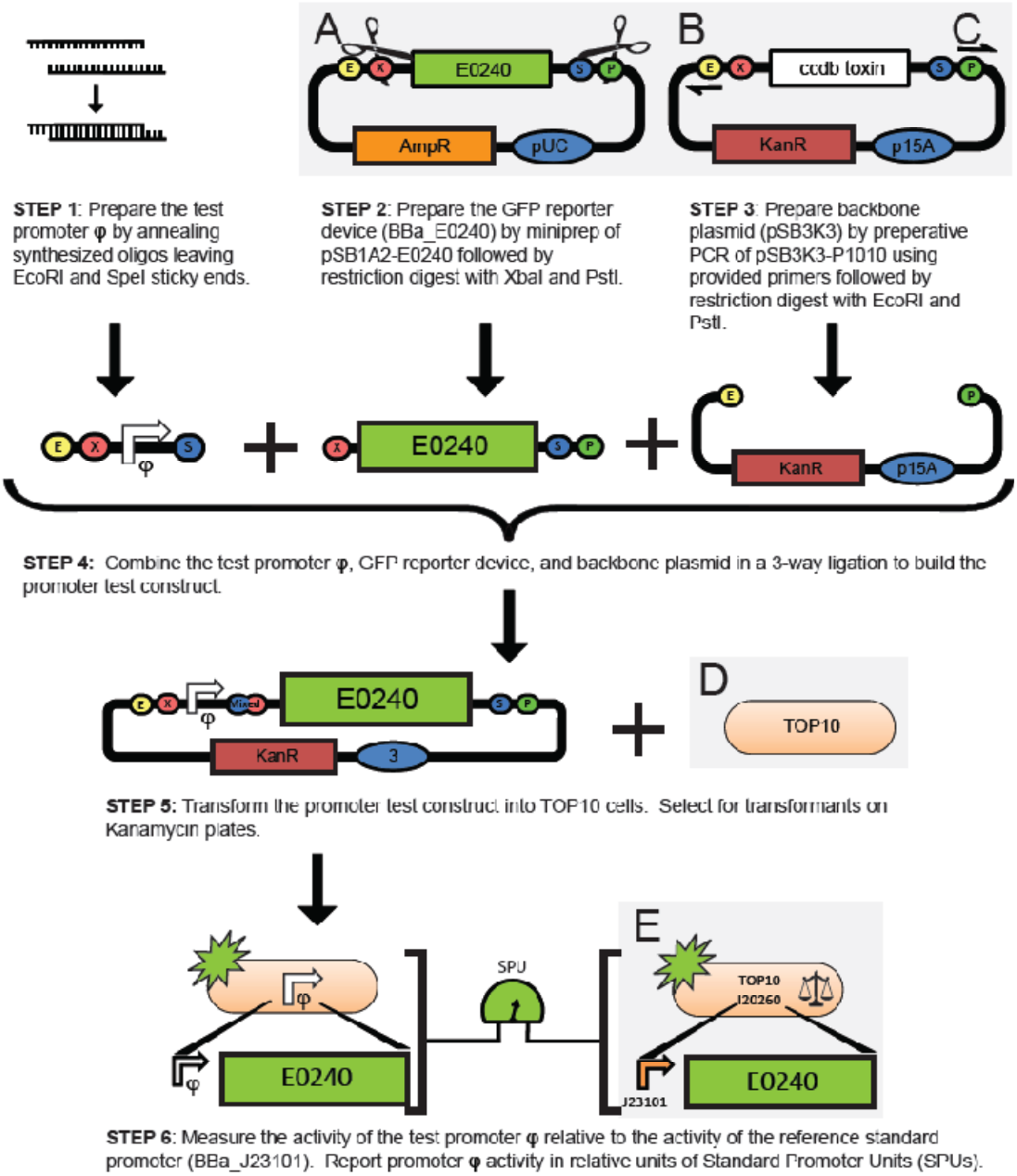


Figure 3-3 Instructions for inserting a promoter into the promoter measurement kit and measuring the promoter activity in Standard Promoter Units (SPUs).

The components marked A-E and shaded in gray are included with the kit: **(A)** pSB1A2-E0240 is provided as purified DNA. BBa_E0240 is the GFP reporter device (containing an RBS, GFP coding region, and transcription terminator). **(B)** pSB3K3-P1010 is provided as purified DNA to serve as template for a preparative PCR reaction. **(C)** DNA primers are

provided for use in a preparative PCR reaction to produce backbone vector (pSB3K3). **(D)** TOP10 cells are provided to serve as the standard characterization and construction strain. **(E)** pSB3K3-I20260 is provided as a plasmid in TOP10. BBa_I20260 contains the promoter reference standard, BBa_J23101, upstream of the GFP measurement device. The instructions for using the RBS measurement kit are similar to the instructions for the promoter measurement kit (Appendix D Figure 1).

The promoter measurement kit contains measurement "tools" such as a green fluorescent protein (GFP) reporter device (BBa_E0240), backbone plasmid (pSB3K3), and *E.coli* strain (TOP10) (Table 3-1). We also designate a specific promoter (BBa_J23101) and RBS (BBa_B0032) as reference standards. This promoter and RBS pair were inserted upstream of GFP and are included in the kits as the reference standard construct (BBa_I20260). In order to measure the activity of a user-specified promoter, kit users assemble the user-specified promoter upstream of the GFP reporter device and insert this combined part into the backbone plasmid to form the promoter test construct. The process for inserting a promoter upstream of the GFP reporter device is based on three-antibiotic BioBrick standard assembly [30], and is outlined in the instructions included with the kit (Figure 3-3). Similar instructions are included in the RBS measurement kit for inserting an RBS upstream of GFP (Figure D-1).

Following assembly, the promoter test construct is transformed into *E.coli* TOP10 cells and these cells are grown along with TOP10 cells containing the reference standard construct. The activity of the user-specified promoter is measured indirectly based on GFP fluorescence in exponential phase and then reported relative to the activity of the reference standard promoter that is measured in parallel by identical procedures (Appendix B). This

relative measurement of activity enables independent researchers to calibrate their measurements and report them in newly defined, comparable units: standard promoter units (SPUs) and standard RBS units (SRUs). Moreover, we have adopted a simple model relating the activity of a promoter or RBS to the GFP synthesis rate from the promoter or RBS test construct. We use this model to demonstrate that normalizing to a reference standard can account for variability in the reaction rates associated with GFP expression and fluorescence (for instance, GFP maturation rate) that might otherwise lead to inconsistent part measurement across varying culture conditions.

To demonstrate the utility of the measurement kits we characterized the activities of seven promoters and five RBSs. Additionally, we characterized four promoters via different measurement procedures to explore whether the promoter reference standard could help to account for variation in measurement procedures that might be expected when experiments are repeated on different days or by different laboratories. Finally, we distributed a set of four promoters to six independent laboratories to obtain a practical estimate of the variation expected in part measurement between independent groups using the kits.

3.3 Results

3.3.1 Measurement kit design & construction

We had four goals in designing the measurement kits. First, we wanted the DNA manipulation necessary for assembling the promoter or RBS test construct to be simple and reliable so that the kits could be adopted as widely as possible. Second, we wanted to choose

a reporter of promoter activity and RBS activity that could be detected under a wide range of possible measurement conditions. For example, researchers might want to characterize parts at different cellular growth phases, under different culture conditions, with different instruments, or at different resolutions (for example, single-cell or in bulk culture). Third, we wanted reference standards for promoter activity and RBS activity that were similar to the activities of other promoters and RBSs commonly used in engineered biological systems so that our measurement kits would be relevant to many biological engineers. Fourth, we wanted the characterization of a promoter or RBS via the measurement kits to provide an accurate estimate of the performance of the promoter or RBS when reused in any engineered system built using standard BioBrick DNA assembly methods.

To meet our first design goal we needed a reliable method to insert a promoter or RBS upstream of GFP in our backbone plasmid, pSB3K3. We designed the DNA parts included in the kits (BBa_E0240, BBa_I13401, and pSB3K3) for assembly via three-antibiotic BioBrick standard assembly methods [30]. These assembly methods have proven simple enough to be reliably implemented by many first-time researchers in the annual International Genetically Engineered Machines (iGEM) Competition [7]. We chose *E.coli* TOP10 (Invitrogen) as the testing strain included in the kits due to the high transformation efficiency of the strain.

To meet our second design goal we needed a robust reporter of gene expression. We chose a fast-folding, enhanced-fluorescence green fluorescent protein (GFP) variant (*gfp*mut3*) as a reporter as it can be measured via many methods [64, 65].

To meet our third design goal we chose as reference standards a 35bp constitutive promoter (BBa_J23101) and a 13bp RBS (BBa_B0032) that are similar in activity to promoters and RBSs that have been previously used in a number of engineered biological systems [22, 38, 66]. For example, the commonly used $P_{Ltet-O1}$ and P_{Lac-O1} promoters [67] have a measured activity within 10% of the activity of the reference standard promoter, while the reference standard RBS is the middle-strength RBS from a collection of RBSs that have been used previously to tune gene expression levels [63, 68].

Meeting our fourth design goal, ensuring that a part characterized in one context behaves equivalently when placed in a different context, is a central challenge in biological engineering [49]. For example, the DNA sequence adjacent to promoters and RBSs is known to affect their activities [69-71]. The BioBrick assembly standard begins to address the problem of insulating parts from the effects of adjacent sequences by including 8 bp "BioBrick standard junctions" between combined parts, although these junctions can create additional challenges such as causing a frameshift when assembling two protein coding regions. Another challenge is that the BioBrick assembly standard only allows insertion of a new part at the beginning or end of a string of parts, which can slow the process of inserting a new RBS between the promoter and the GFP reporter device. Thus, we evaluated an alternate cloning scheme for assembling the RBS test construct. This alternate scheme allowed for easier insertion of the RBS than the standard BioBrick assembly but created junction sequences differing slightly from the standard BioBrick junction (Appendix D). After testing, we rejected the alternate design because the measured RBS activities when

flanked by non-standard junctions were not equivalent to the activities measured when the same parts were flanked by standard BioBrick junctions (not shown). Thus, we chose to use three-antibiotic BioBrick standard assembly [30] in our final design to ensure that parts inserted in the promoter or RBS test construct would be flanked by the same junction sequences as the parts would be when inserted into any engineered system built via BioBrick standard assembly methods.

3.3.2 Definitions and models for promoter and RBS activity

In order to quantitatively characterize promoters and RBSs we need a clear definition of each part and the functional characteristics to be measured. We define a BioBrick promoter as any DNA sequence meeting the BioBrick assembly standard [40] that produces an output of RNA polymerases that transcribe the DNA downstream of the promoter. The rate of polymerases exiting a promoter to synthesize mRNA (rate of successful mRNA initiation events or polymerases per second, PoPS) specifies the activity of the promoter [72]. Similarly, we define a BioBrick RBS as any RNA sequence that produces an output of ribosomes that translate the RNA downstream of the RBS. The rate of ribosomes exiting the RBS to synthesize protein specifies the activity of the RBS. In order to calibrate measurements, users of the measurement kits should report promoter and RBS activities in relative units (Standard Promoter Units, SPUs and Standard RBS Units, SRUs). These units are defined as the ratio of the steady-state activity of a user-specified promoter or RBS (for

instance, promoter φ or RBS φ) to the steady-state activity of the reference standard promoter (BBa_J23101) or RBS (BBa_B0032):

$$Activity\ of\ Promoter\ \varphi\ (SPUs) = \frac{PoPS_{\varphi}^{SS}}{PoPS_{J23101}^{SS}} \qquad (Eq.\ 3.1)$$

where $PoPS^{SS}$ is the steady-state rate of successful mRNA initiation events in absolute units of mRNA per second per DNA copy of the promoter and the subscripts φ or BBa_J23101 refer to the rates associated with a user-specified promoter (φ) or rates associated with the reference standard promoter, respectively. We adopted a previously described ordinary differential equation (ODE) model of GFP expression from a constitutive promoter [49, 73] to model the GFP expression from the promoter test construct and the reference standard construct. This model relates GFP synthesis rate per cell to the rates associated with the various steps in the production of mature GFP and to the promoter activity measured in absolute molecular units, such as PoPS. We can evaluate this ODE model at steady-state (Appendix D) in order to determine the rate of successful mRNA initiation events per DNA copy of the promoter ($PoPS^{SS}$) from the mature GFP synthesis rate per cell (S_{cell}^{SS}):

$$PoPS^{SS} = \frac{\gamma_M(a+\gamma_I)S_{cell}^{SS}}{\rho a n} \qquad (Eq.\ 3.2)$$

where γ_M is the mRNA degradation rate, a is the GFP maturation rate, γ_I is the degradation rate of immature GFP, ρ is the translation rate of immature GFP from mRNA, and n is the number of copies of the promoter in the cell. To estimate promoter activity from GFP synthesis rates it is necessary to determine each of these five parameters. In practice measuring all five parameters for each new promoter is challenging and researchers have

often turned to previously published values of rates in parameterizing their models [62]. However, the published rates may be inaccurate descriptions of the system being modeled as they are typically measured under different culture conditions and with a different DNA construct expressing GFP. Instead we can account for some of these rates by making use of the reference standard promoter included in the promoter measurement kit. Rather than making the assumption that the parameters $(\gamma_M, a, \gamma_I, \rho, n)$ describing our promoter test construct are equivalent to previously published literature values, we instead assume that the parameters are equivalent between two promoters inserted in identical DNA constructs. Thus, by combining Eq. 3.1 and Eq. 3.2, we can calculate promoter strength in relative units of SPU and eliminate some of these rates via cancellation of terms:

$$Activity\ of\ Promoter\ \varphi\ (SPU) = \frac{\frac{\gamma_{M,\varphi}(a_\varphi + \gamma_{I,\varphi})S^{SS}_{cell,\varphi}}{\rho_\varphi a_\varphi n_\varphi}}{\frac{\gamma_{M,J23101}(a_{J23101} + \gamma_{I,J23101})S^{SS}_{cell,J23101}}{\rho_{J23101} a_{J23101} n_{J23101}}} \quad (Eq\ 3.3)$$

We can make additional assumptions to further simplify Eq. 3. First, we can assume that GFP expressed from either the test promoter φ or the reference standard promoter has an equivalent maturation rate ($a_\varphi = a_{J23101} = a$) since the two promoters are measured under the same culture conditions. Second, since both promoters are carried on the same backbone plasmid, we assume that each promoter is at the same average copy number ($n_\varphi = n_{J23101}$). There are reported cases of promoter activity influencing the copy number of plasmids due to polymerases transcribing through the plasmid origin of replication [74], however the transcription terminator (BBa_B0015) downstream of GFP in the promoter test construct as

well as the transcription terminators flanking the BioBrick cloning site [30] should largely prevent differences in promoter activity from causing differences in plasmid copy number. Third, since the promoters tested here have been standardized to have identical transcription initiation sites (predicted) and identical sequences downstream of the initiation site (Appendix D) we expect that each promoter produces the same mRNA sequence [75]. If the transcribed mRNAs are identical then we can assume that the mRNA degradation rates are equivalent ($\gamma_{M,\varphi} = \gamma_{M,J23101}$) and the translation rates of immature GFP from mRNA are also equivalent ($\rho_{\varphi} = \rho_{J23101}$). mRNA degradation is also a function of dilution due to cellular growth, however the dilution rate is negligible relative to typical rates of active mRNA degradation in *E.coli* [76]. Lastly, we can assume that immature GFP is stable so that protein degradation is negligible compared to dilution due to cellular growth ($\gamma_{I,\varphi} = \mu_{\varphi}$ and $\gamma_{I,J23101} = \mu_{J23101}$, where μ is the cellular growth rate). Following the above assumptions, Eq. 3.3 can be simplified to:

$$Activity\ of\ Promoter\ \varphi\ (SPU) = \frac{(a + \mu_{\varphi})S^{SS}_{cell,\varphi}}{(a + \mu_{J23101})S^{SS}_{cell,J23101}} \qquad (Eq. 3.4)$$

This equation can be simplified further by noting that:

$$if\ |\mu_{\varphi} - \mu_{J23101}| \ll a\ \ then\ \ \frac{(a + \mu_{\varphi})}{(a + \mu_{J23101})} \approx 1 \qquad (Eq. 3.5)$$

For example, we measured the growth rates of cells in our experiments to determine if the difference between the growth rates of cells containing the promoter test construct (μ_{φ}) and cells containing the reference standard construct (μ_{J23101}) is negligible compared to the

maturation rate of GFP (that is,$\left| \mu_{\varphi} - \mu_{J23101} \right| \ll a$). The cellular growth rates varied depending on the promoter being tested as well as on the experimental conditions: the fastest growth rate was observed in cells containing the BBa_J23113 promoter test construct grown in M9+glucose (μ = 2.5E-4 sec^{-1}) and the slowest growth rate was observed in cells containing the BBa_R0040 promoter test construct grown in M9+glycerol (μ=1.39E-4 sec^{-1}). The maturation rate of the GFP variant used in the GFP reporter device (BBa_E0040) has been previously measured as a =1.8E-3 sec^{-1} [49]. Based on the worst-case assumption that the cells containing the promoter test construct are the fastest growing cells(μ_{φ} =2.5E-4sec^{-1}) and the cells containing the reference standard construct are the slowest growing cells (μ_{J23101} = 1.39E-4 sec^{-1}) then:

$$\frac{(a + \mu_{\varphi})}{(a + \mu_{J23101})} = 1.06 \approx 1 \qquad (Eq. 3.6)$$

Therefore, if we assume that the difference between the growth rates of cells containing the promoter test construct (μ_{φ}) and cells containing the reference standard construct (μ_{J23101}) is negligible compared to the maturation rate of GFP then Eq. 4 can be combined with Eq. 5 yielding:

$$Activity\ of\ Promoter\ \varphi\ (SPU) = \frac{S^{SS}_{cell,\varphi}}{S^{SS}_{cell,J23101}} \qquad (Eq. 3.7)$$

Thus, by reporting promoter activity relative to a reference standard promoter (BBa_J23101) and choosing promoters with identical transcription initiation sites and identical sequences downstream of the initiation sites, researchers can report measured promoter activities in

compatible units without having to independently measure the GFP maturation rate, mRNA degradation rate, protein production rate, or plasmid copy number for their specific experimental setup. The assumptions of equivalence in the parameters associated with GFP synthesis ($\gamma_M, a, \gamma_I, \rho, n$) between the reference standard promoter and the user-specified promoter will only hold across an as-yet unknown range of operating conditions and promoter activities; the specifics of the acceptable operating conditions will be better defined as biological engineers make use of the measurement kits.

We derived similar equations for RBS activity defined in terms of SRUs (Appendix D Eqs. D.13-D.17), yielding the following equation:

$$Activity\ of\ RBS\ \varphi\ (SRU) = \frac{S^{SS}_{cell,\varphi}}{S^{SS}_{cell,B0032}} \qquad (Eq.\ 3.8)$$

While most of the assumptions made for the promoter measurement kit can be applied to the RBS measurement kit, the assumption that the degradation rate of mRNA produced by the RBS test construct ($\gamma_{M,\varphi}$) is equivalent to the degradation rate of mRNA produced by the reference standard construct ($\gamma_{M,B0032}$) is not fully justified. For example, our previous assumption of equivalence in mRNA degradation rates between the promoter test construct and the reference standard construct was based on these two promoters producing identical mRNAs; however it is impossible for the RBS test construct and the reference standard construct to produce identical mRNAs due to the different RBS sequences present in the different mRNAs. We begin to address this issue here by choosing RBSs with similar lengths and sequences in order to minimize the differences in the mRNA and thus support

the assumption of equivalent mRNA degradation rates ($\gamma_{M,\varphi} = \gamma_{M,B0032}$). However, further work is needed to better predict mRNA degradation rates from sequences or to design flanking sequences that might fix mRNA degradation rates at a predictable value [33].

3.3.3 Demonstration of Measurement kit utility

We measured the activities of seven promoters and five RBSs (Figure 3-4) obtained from the Registry of Standard Biological Parts [8]. These promoters included members of a constitutive promoter library (BBa_J23100 – BBa_J23119, constructed by JC Anderson) as well as commonly used Tet repressor (BBa_R0040) and Lac repressor (BBa_R0011) regulated promoters [67]. The regulated promoters were tested in the absence of their cognate repressor proteins. The set of five RBSs had been used previously to tune protein expression [63]. This set of promoters and RBSs also serves as a 12-member library of promoter-RBS pairs providing a greater than 100-fold range in gene expression levels. Such libraries of characterized promoters and RBSs of varying activities have been shown to be valuable to researchers for tuning biochemical networks to optimize the synthesis of products of interest [62, 68].

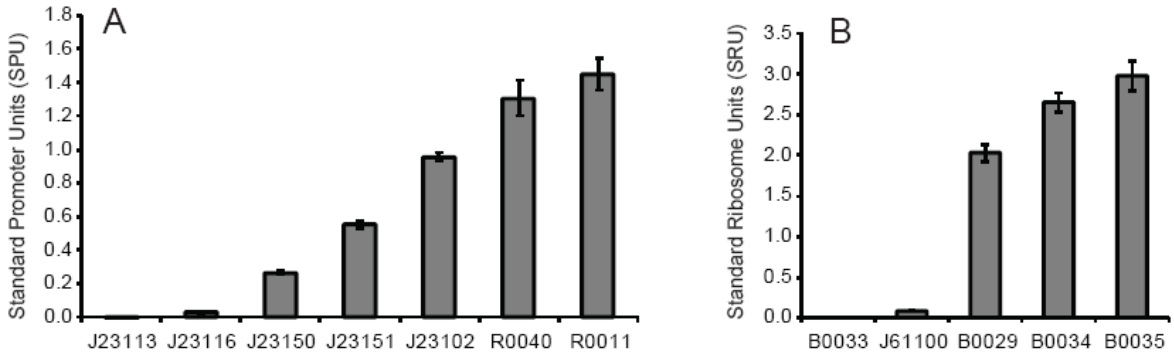


Figure 3-4 Measurement of the activity of a test set of (A) promoters and (B) RBSs using the measurement kits.

The five promoters labeled J23### are from a constitutive promoter library and R0040 and R0011 are tet- and lac-repressible promoters, respectively. The activity of the promoters was measured in standard promoter units (SPUs) and the activity of the RBSs was measured in standard RBS units (SRUs). This collection of promoter-RBS pairs may itself be useful for tuning gene expression in engineered systems. The error bars represent the 95% confidence interval of the mean based on nine replicates.

We measured the promoter and RBS activities by calculating the steady-state GFP synthesis rates (Appendix B) and converting these rates to SPUs and SRUs. Nine independent clones were characterized across three separate experimental runs for each promoter tested. The promoters ranged in activity from 0.026 ± 0.003 to 1.45 ± 0.095 SPUs (uncertainties throughout represent 95% confidence interval of the mean). Ribosome binding sites ranged in activity from 0.093 ± 0.006 to 2.99 ± 0.188 SRUs. The GFP expression levels from one promoter (BBa_J23113) and one RBS (BBa_B0033) were statistically equivalent within measurement error to the expression level of the negative control (TOP10).

3.3.4 Variability due to measurement procedures

Maintaining a consistent measurement procedure across independent laboratories can be challenging [50]. Furthermore, we do not know which aspects of measurement procedures are most likely to influence the assumptions that allow for calculation of SPUs and SRUs from GFP expression data. To begin to explore the effect of different measurement procedures on promoter activity as measured with the kit, we characterized four BioBrick promoters (BBa_J23113, BBa_J23150, BBa_J23151, and BBa_J23102) using four different measurement procedures (Figure 3-5). We varied the measurement procedures by changing: (1) the instrument used for measurement (multi-well fluorimeter or flow cytometer), (2) the culture media (M9 with glucose or M9 with glycerol), and (3) the culture growth conditions (tubes on a roller or 96-well plate with shaking).

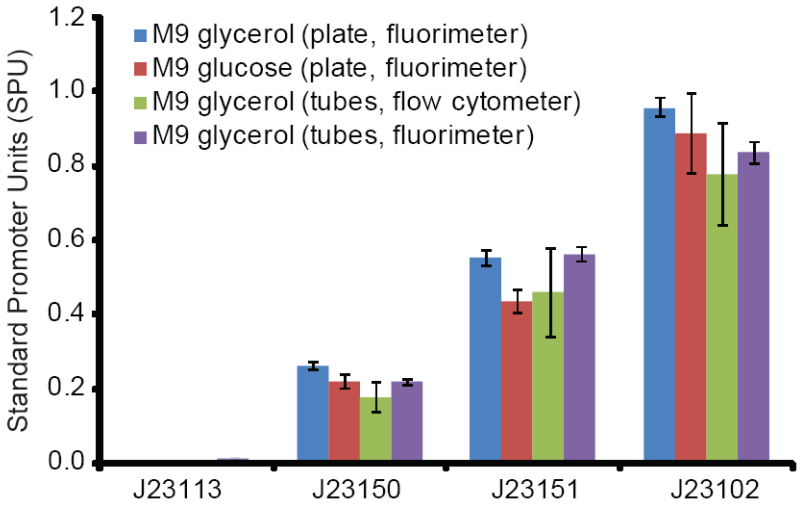


Figure 3-5 Measurement of the activity of a set of four promoters using four different measurement procedures

Blue: cells grown in M9+glycerol in a 96-well plate with GFP synthesis rate measured by a fluorimeter, *red:* cells grown in M9+glucose in a 96-well plate with GFP synthesis rate measured with a fluorimeter, *green:* cells grown in M9+glycerol in culture tubes with GFP concentration measured on a flow cytometer, *purple:* cells grown in M9+glycerol in culture tubes with GFP concentration measured with a fluorimeter. The error bars represent the 95% confidence interval of the mean based on either nine replicates (blue and purple) or three replicates (red and green). The activity of BBa_J23113 was equivalent to the negative control within error. The coefficient of variation of the means of the other three promoters across the four measurement procedures were 0.163, 0.128, and 0.099 for promoters BBa_J23150, BBa_J23151, and BBa_J23102, respectively.

The method we used to quantify fluorescence varied depending on the instrument used to make our measurements. We collected bulk fluorescence and absorbance measurements from the multi-well fluorimeter and geometric means of the population per-cell fluorescence measured by the flow cytometer (Appendix B). We used the multi-well fluorimeter to capture absorbance and fluorescence levels over time in order to calculate GFP synthesis rate per cell. We used a flow cytometer to capture a single time point, allowing for

calculation of GFP concentration per cell but not GFP synthesis rate. In order to make use of GFP concentration per cell we expanded our previously described model (Eq. 3.1-3.7) to include cellular growth rate and GFP concentration (Appendix B). To parameterize this model we measured the growth rates of cells containing the promoter test constructs with each of the individual promoters (Appendix D Table 3). We compared the promoter activity in SPUs across the four measurement procedures for each promoter (Fig. 4) and found coefficients of variation of 0.163, 0.128, and 0.099 for promoters BBa_J23150, BBa_J23151, and BBa_J23102, respectively. As before, the activity of BBa_J23113 was indistinguishable from the negative control within the error of our measurements. The consistency of the measured activity of the promoters under a variety of conditions suggested that measurements taken via the kit might be robust to the minor variations that are expected when following an identical procedure on different days or in different laboratories.

3.3.5 Inter-laboratory variation

To make a practical estimate of the variation in the measured activity of promoters across laboratories we distributed four strains containing promoter test constructs (BBa_J23113, BBa_J23150, BBa_J23151, and BBa_J23102) to researchers in six other laboratories. Each researcher independently measured the activity of the four promoters using the same procedure: (1) three independent cultures were grown from single colonies for each of the four promoters, (2) cells were collected in exponential phase, (3) the GFP concentration per cell was measured using a flow cytometer, (4) the flow cytometer data was

gated based on forward and side scatter and the negative control, and (5) the geometric mean of the per cell fluorescence in the population was reported for each culture (Appendix B). No special effort was made to standardize the flow cytometers used or the machine settings beyond asking researchers to use typical settings for measuring GFP and by providing each lab with an example plot to guide gating of the flow cytometer data based on forward scatter, side scatter, and fluorescence [49]. We compared the promoter activity in SPUs across the six laboratories (Figure 3-6) and found coefficients of variation of 0.164, 0.118, and 0.059 for promoters BBa_J23150, BBa_J23151, and BBa_J23102, respectively. As before, the activity of BBa_J23112 was indistinguishable from the negative control within the error of our measurements for all but one laboratory. As expected, there were slight differences in how the protocol was conducted in each laboratory, such as different culture conditions (rollers or shakers) and growth time, however the reported promoter activities were fairly insensitive to these differences.

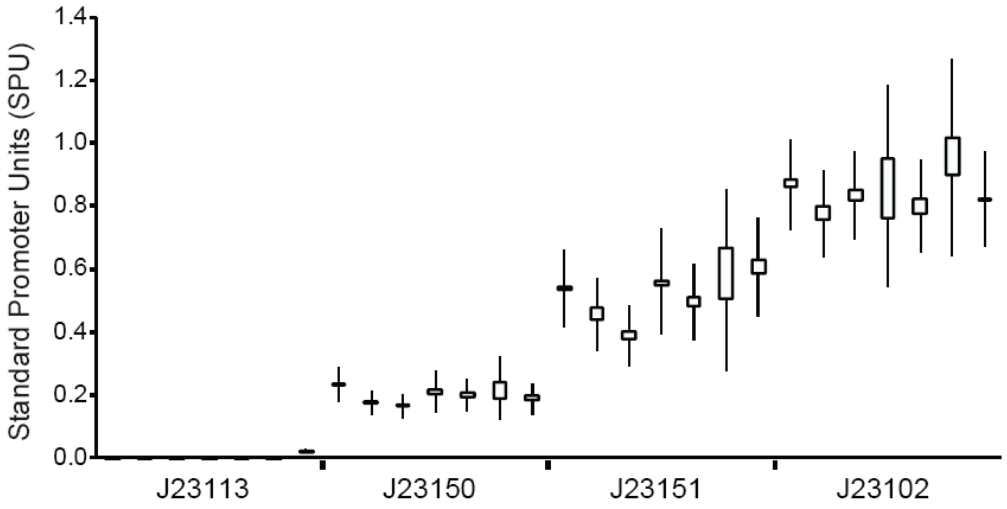


Figure 3-6 Comparison of promoter activities measured by researchers in seven independent laboratories using the measurement kit.

Each laboratory followed the same procedure and the measured activity is based on GFP concentration measured via a flow cytometer. Measurements were taken in triplicate; the boxes show the range of measured activities and the lines show the 95% confidence interval of the mean. The activity of BBa_J23113 was equivalent to the negative control within error for all but one of the laboratories. The coefficient of variation of the means of the other three promoters across the four measurement procedures were 0.164, 0.118, and 0.059 for promoters BBa_J23150, BBa_J23151, and BBa_J23102, respectively.

3.4 Discussion

We designed, built, and tested kits for making standard measurements of promoter and RBS activity. The kits include reference standards that enable researchers to report promoter or RBS activity in common units, Standard Promoter Units (SPUs) or Standard RBS Units (SRUs), respectively. We demonstrated the functionality of the kits by measuring the activity of sets of promoters and RBS parts. To explore the effect of different characterization procedures on measured promoter activity we used four different procedures

to characterize four promoters and found that the activity of each promoter in SPUs was consistent across the procedures with an average coefficient of variation of 0.127. Finally, we distributed four promoters to researchers in six independent laboratories and found an average coefficient of variation of 0.113 in the reported activity of each promoter across the laboratories. The relatively high degree of reproducibility of measurements between laboratories is encouraging for future work to improve and develop standard measurement kits.

3.4.1 Suggested improvements

The measurement kits are an early attempt at providing shared measurement tools and reference standards to a community of biological engineers. We expect that this community will improve or replace various kit components in order to improve the accuracy, ease of use, reliability or sensitivity of the kits. Many improvements to the kits could be imagined, and here we list six that might be of immediate use: (1) locating the promoter test construct on the chromosome rather than integrating it into a plasmid, (2) more sensitive reporters for gene expression, (3) multiple reporters of gene expression, (4) constitutive expression of a second reporter in the test strain to provide a control for extrinsic noise, (5) alternate reference standards for promoters or RBSs with lower variability and activity, and (6) increasing the number of validated strains for the kits.

The use of plasmid-based reference standards can lead to increased variability in reported promoter and RBS activities due to apparent cell to cell variation in plasmid copy

number [77]. Additionally, promoter strength could influence plasmid copy number in unexpected ways, such as by promoters reading through the transcription terminator and effecting plasmid replication by transcribing within the origin of replication [74]. Additionally, we saw differences in growth rate among different strength promoters and RBSs (Appendix D Table 3). These growth rate differences are likely a result of overuse of cellular resource for the synthesis of GFP and make measurement of the promoter or RBS activity more error-prone, as cells growing at different rates can have different numbers of available polymerases or ribosomes [78]. Moving the promoter and RBS test constructs to the genome or a lower-copy plasmid could help to further mitigate both effects.

While GFP has many advantages as a reporter of gene expression, future versions of the kits might alter the reporter in four different ways. First, enzymatic reporters such as β-gal or luciferase would have greater sensitivity and facilitate characterization of promoters and RBSs with weak activities. Second, measurement kits with multiple reporters could be used to gauge how sensitive a promoter or RBS is to adjacent genetic sequences by comparing activities across reporters encoded by distinct sequences. Third, reporter genes with degradation tags that accelerate degradation of the reporter protein would be useful for exploring dynamic responses. Fourth, constitutive expression of a second reporter gene in addition to GFP (for example red fluorescent protein or β-gal) from the promoter test construct may provide an internal control for sources of extrinsic noise such as differences in growth rate, variation in plasmid copy number, or fluctuations in concentrations of cellular factors [61].

A better reference standard promoter or RBS might be chosen in the future. For example, cells containing the reference standard construct have a growth rate that is ~ 10% lower than cells lacking the construct. A weaker promoter or RBS might be chosen that has less of an impact on cellular growth rate. Additionally, a future reference standard promoter or RBS might be chosen that has a particularly narrow distribution of promoter activity across the population of cells containing the reference standard construct. A narrow distribution in activity would help to reduce error in calculating SPUs, since the error in the measured activity of the reference standard is propagated to all SPU measurements. As a general rule, reference standards should not be changed since promoter activity measured relative to a new reference standard will not be in compatible units to promoter activity measured relative to an older reference standard (below). However, since the work here presents first generation measurement kits we expect that significant improvements could be made to the reference standards described here that might justify updating the measurement standards themselves.

The kits include TOP10 as the standard testing strain in order to ensure a common genetic background when characterizing parts. It is unknown how differences in genetic background would affect the assumptions in our model for calculating SPUs and SRUs. However, many differences, such as differences in the number of ribosomes or polymerases, might affect the reference standard promoter or RBS and the user-specified promoters or RBSs equivalently. Thus, normalizing to a reference standard may account for differences between many common cloning strains. A follow-on experiment evaluating the performance

of the measurement kits in different genetic backgrounds might expand the acceptable list of testing strains and provide users of the kits with the flexibility to use their strain of choice when characterizing parts.

3.4.2 Standard promoter and RBS definition

In our model relating GFP synthesis to SPUs we assume identical transcription start sites and identical sequences downstream of the transcription starts for all promoters tested. Thus, we expect that the mRNA expressed by each of the promoters tested is identical to the mRNA produced by the reference standard promoter and that we can cancel the mRNA degradation rate and the translation rate of immature GFP from mRNA in order to simplify our model. This critical simplification allows the activity of parts to be reported in comparable units (SPUs) without needing to directly measure mRNA levels. Going forward, a standard promoter definition is needed to ensure that all promoters have a fixed transcription start position and fixed sequence downstream of the transcription start site. All promoters that adhered to such a standard could then be reliably measured using the kits described here or via future kits with alternate reporters of gene expression. However, if such a standard is not adhered to then researchers measuring promoter activity will need to quantify the rate of mRNA degradation for the different mRNAs produced by each non-standard promoter.

Defining a standard RBS element is more challenging, since the mRNA produced by the RBS test construct and the reference standard construct will necessarily be different due

to the different RBS sequences on the mRNA. Future work to better predict mRNA degradation rates from primary sequences or to design flanking sequences that might fix the degradation rates of mRNA at one predictable value [33] will be needed to address this challenge.

3.4.3 Measurement procedures

The measurement kits provide a shared platform for researchers to evaluate different measurement procedures. We deliberately have not advocated a single measurement procedure for characterizing promoters or RBSs via the measurement kits. The choice of the best measurement procedure will be influenced by the particular groups making use of the kits. For example, laboratories without access to equipment for capturing high-throughput single-cell measurements of fluorescence might opt for a bulk fluorescence measurement using a fluorimeter. Other groups might prefer to obtain single-cell measurements using quantitative microscopy or flow cytometry, or to capture a time-course of fluorescence measurements from a growing culture. As different measurement procedures are likely to have merits within different communities we expect that a number of procedures will be established. As an example, for the community of undergraduate iGEM teams using the measurement kits, we have suggested a measurement protocol that can be easily carried out by novice researchers and that only requires two absorbance and bulk fluorescence measurements [47].

3.4.4 Absolute and relative units

The new units specified here, SPUs and SRUs, are relative units relating the activity of a user-specified promoter or RBS to the reference promoter or RBS (BBa_J23101 or BBa_B0032). We can also define promoter activity in "absolute" units defined as the steady-state rate of successful mRNA initiation events per DNA copy of the promoter (represented by $PoPS^{SS}$ in our model). PoPS is challenging to measure directly, however it is essential that researchers be able to convert from SPUs to absolute units as absolute units tie promoter activity to the larger system of related reactions within the cell. For instance, the PoPS rate could be used to make an estimate of the demand placed on available pools of polymerases in the cell by a particular promoter [49]. We have made a rough estimate of the relationship between SPUs and PoPS and found 1.0 SPU ≈ 0.03 PoPS (Appendix D). We also estimated the relationship between SRUs and RBS activity in absolute units (ρ, protein initiation events per second per mRNA) and found 1.0 SRU ≈ 0.04 proteins initiated per second per mRNA (Appendix D).

Should SPUs then be defined in terms of absolute units? No. We are informed by the decisions of previous standards-setting bodies to define units (such as the ohm) based on the properties of unchanging, physical objects rather than by absolute units [57, 58]. For example, we could define 1 SPU in absolute units of 1 PoPS, and then construct a promoter to the best of our current ability that produces 1.0 PoPS and that would then serve as a reference standard. However, as our ability to accurately measure mRNA production improved, we would inevitably discover that our 1 SPU reference standard promoter did not

produce exactly 1.0 PoPS. We would then be forced to redesign the reference standard again and again so that it accurately represented 1 SPU (1.0 PoPS) according to our ever-improving measurement capabilities. However, altering the reference standard is problematic as promoters measured relative to the new standard would be "out of sync" with promoters measured against prior standards. On the other hand, if instead of using absolute units, we define SPUs as the PoPS produced by a particular promoter (for instance, BBa_J23101), then our reference standard remains unchanging even as improved methods for measuring the absolute PoPS rate of BBa_J23101 are developed. In other words, the reference standard for 1 SPU will not change while the relationship between 1 SPU and PoPS will evolve over time given more accurate measurements. In this arrangement, a promoter with activity measured as 2 SPUs will always have an activity of 2 SPUs, even if over time 2 SPUs transitions from being equivalent to 2.0 PoPS to 2.1 PoPS. Thus, as we define new units of measurement for describing the activities of other types of biological parts it will be preferable to choose units that are defined by the activities of particular reference standards in order to ensure that measurements reported in such units will remain compatible over time.

3.4.5 Distribution, use, and improvement of kits

Shared measurement tools and reference standards are only useful if they are adopted within a community. To facilitate distribution of the measurement kits we have worked with the Registry of Standard Biological Parts to include the kits in the annual distribution of

BioBrick parts. To facilitate the reporting and sharing of measurements of promoter and RBS activities we have set up a website (http://partsregistry.org/measurement). This website also contains instructions for use of the kits and lists of previously characterized parts (including the 12 promoters and RBSs described here). Finally, to enable discussion of proposed improvements to the kits and reference standards, and also the development of new kits and reference standards, we are supporting an open discussion of technical standards in synthetic biology (http://biobricks.org/standards).

3.5 Conclusions

Standard tools, techniques, and units for measurement are needed for a distributed community of biological engineers to independently characterize and share biological parts. We have developed first-generation measurement kits for BioBrick promoters and RBSs, and are distributing these kits to users of the MIT Registry of Standard Biological Parts. Having demonstrated the feasibility and ease of use of the kits, we hope to encourage a community of users to adopt these measurement tools and reference standards in order to characterize parts in comparable, common units (SPUs and SRUs). We expect that the shared experiences of biological engineers using common tools and standards will help to identify new engineering challenges in improving the reliability and robustness of standard biological parts.

Chapter 4. Sortostat: An integrated microchemostat and optical cell sorting system

[The work presented in this chapter was was carried out in collaboration with Josh Michener and Bryan Hernandez]

4.1 Summary

I describe a new microfluidic device, the Sortostat, that integrates a cell sorting chamber with a previously published microscope-mounted microfluidic chemostat. Researchers can use the Sortostat to apply morphological, time-varying, or other complex selective pressure to cells in continuous culture. Selection is based on automated processing of images captured on a microscope and thus unique selective pressures can be applied with the Sortostat based on any information that can be derived from microscope images of cells. Tools such as the Sortostat will be needed to apply directed evolution approaches to screen for more complicated engineered biological systems.

Here, I describe the design and construction of the Sortostat microfluidic chip as well as the automated platform used to operate the chip. I validate expected mathematical distributions that describe aspects of the device performance. I demonstrate the application of unique selective pressures with the Sortostat by independently varying the densities of two competing subpopulations of *E.coli* engineered to express either cyan fluorescent protein (CFP) or yellow fluorescent protein (YFP). Furthermore, I demonstrate that the Sortostat can

be used to select in favor of slower-growing subpopulations and stably maintain these disadvantaged bacteria in the Sortostat for many hours.

4.2 Introduction

Researchers have made use of directed evolution to successfully engineer biological parts such as enzymes [2], transcriptional promoters [3], and fluorescent proteins [4] by screening libraries of variants for mutants with desired phenotypes. The application of directed evolution to more complicated engineered biological systems (Chapter 1) will require new tools for applying complicated screens and selections. For instance, biological engineers might screen for unique morphological characteristics [79, 80], dynamic system performance [22], response to time-varying stimuli [81], or many other characteristics. In this chapter I describe a new device, the Sortostat, that integrates a cell sorting chamber with a previously published microscope-mounted microfluidic chemostat [82]. Researchers can use the Sortostat to apply complex, time-varying selective pressure to cells in continuous culture.

The Sortostat is an extension of a previous design of a microfluidic chemostat (a microchemostat, Figure 4-1) by Frederick Balagadde [82, 83]. A chemostat is a bioreactor that enables continuous culture of a bacterial population by periodically replacing a fraction of the culture volume with fresh media [84, 85]. Chemostats are useful for the characterization of biological systems as they provide a consistent culture environment over long periods of time [45, 84, 85]. After an initial period of growth without nutrient

limitation, the cells in the chemostat eventually deplete an essential nutrient such that their growth rate is limited. From this point onward, the cellular growth rate becomes a function of the input rate of this resource (in other words, the flow rate of new media entering the chemostat). By controlling the input flow rate of fresh media and thus the output flow rate of cells and used media, a researcher can externally set the growth rate of cells in the reactor (Figure 4-1A).

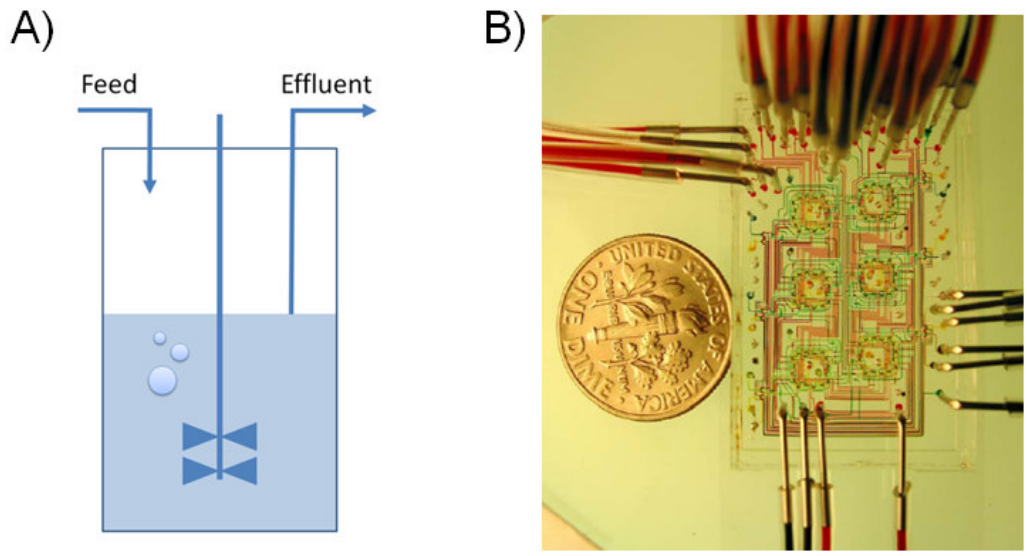


Figure 4-1 A schematic of chemostat and a picture of the microchemostat.

(A) A schematic of a chemostat showing the feed of fresh media entering a stirred reactor containing the cell culture and the effluent (media and cells) leaving the reactor. The inflow and outflow rates are equivalent so the reactor volume remains constant. (B) A single microfluidic chip contains 6 microchemostats that are run in parallel. Food dye has been loaded into the fluid lines to make them easier to visualize. Chemostat image from Wikipedia [86]. Microchemostat image by Frederick Balagadde [83].

Traditional benchtop chemostats have previously been used for the evolution of biological systems. For instance, chemostats have been used to evolve industrial microorganisms with faster growth rates or nutrient uptake rates [87]. Furthermore, multiple strains can be co-cultured in a chemostat in order to select for the strain that is best adapted to a particular set of media conditions [88, 89]. While benchtop chemostats have proven useful for a number of applications, they are challenging to operate reliably due to frequent contamination of the fresh media feed lines, consumption of large amounts of media over long-term culture experiments, and the growth of biofilms that cannot be diluted out of the chemostat and thus interfere with the steady-state growth rate of the overall cell culture [90, 91]. Microchemostats address these challenges. Due to the small reactor volume (16 nanoliters) microchemostats dramatically reduce the media consumption needed to run a long-term experiment. Additionally, the precise control over fluid made possible in the low Reynolds number regime of the microfluidic channels reduces the opportunities for contamination. Although the high surface area to volume ratio in the microchemostat exacerbates the issue of biofilm growth, inclusion of non-adhesive surface coatings such as Bovine Serum Albumin (BSA) in the media or the active cleaning of channels with lysis buffer before refilling with fresh media can be used to prevent the formation of biofilms in the microchemostat [83].

The benchtop chemostat and the microchemostat are both limited to selecting for mutants that improve the growth rate of cells. In order to enable more complicated selections unrelated to cellular growth rates, I extended the design of the microchemostat by integrating

a cell sorting chamber. A small collection of cells can be isolated in the sorting chamber ($1/100^{th}$ of the reactor volume) and, via automated microscopy and image processing, these cells can be sent to the waste or returned to continue growing in the reactor. Thus, in the Sortostat the removal of a fraction of the culture is no longer a random process, but rather a directed process that can be guided by any information that can be gleaned from processing an image of cells held in the sorting chamber. As an example of the utility of the Sortostat, researchers studying diatoms previously imagined a similar device (a 'compustat') to enable the selection of unique morphologies [79, 80].

Here, I describe the design and construction of the Sortostat, as well as validate expected mathematical distributions that describe specific aspects of the device performance. I compare these theoretical models to experimental results, and demonstrate the application of unique selective pressures by independently varying the cell densities of two competing subpopulations of *E.coli* engineered to express either cyan fluorescent protein (CFP) or yellow fluorescent protein (YFP). Furthermore, we demonstrate that we can select in favor of slower-growing subpopulations and stably maintain these disadvantaged bacteria in the Sortostat for many generations.

4.3 Results and Discussion

4.3.1 Design and operation of the microchemostat

Here, I review the details of the microchemostat relevant to the design of the Sortostat, however more detailed descriptions are available elsewhere [82, 83]. The

microchemostat was fabricated from silicone elastomer polydmethylsiloxane (PDMS) using multi-layer soft lithography [92]. Each PDMS chip contained six microchemostats that are run in parallel (Figure 4-1), and are made up of well established microfluidic components such as channels, valves (Figure 4-2), and peristaltic pumps [59, 92]. The microchemostat is an integrated system consisting of a growth chamber loop (10µm high, 140µm wide, and 11.5mm in circumference), micromechanical valves and channels for moving fluids throughout the reactor, and an integrated peristaltic pump for mixing cells and media (Figure 4-3). The growth chamber loop is made up of 16 addressable segments.

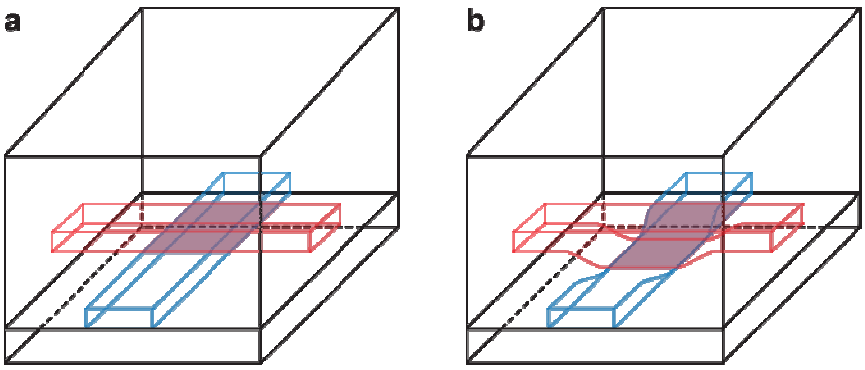


Figure 4-2 Cutaway view of a "push down" PDMS microfluidic valve

The blue channel represents a fluid line that would contain cells and media in the microchemostat. The red channel represents a "push-down" control line, it is filled with water and that can be pressurized or de-pressurized via computer control. (A) When the control channel is unpressurized the valve is "open" and cells and media can flow freely. (B) When the control valve is pressurized the pressure deforms the PDMS between the control line and fluid line and seals the fluid line resulting in a "closed" valve that prevents fluid flow. "Push-up" valves work by a similar principle except they are located below the fluid lines (grey layer in figure). When they are pressurized they deform the PDMS upwards and seal the fluid line by the same mechanism shown here. Image courtesy of Ty Thomson [81].

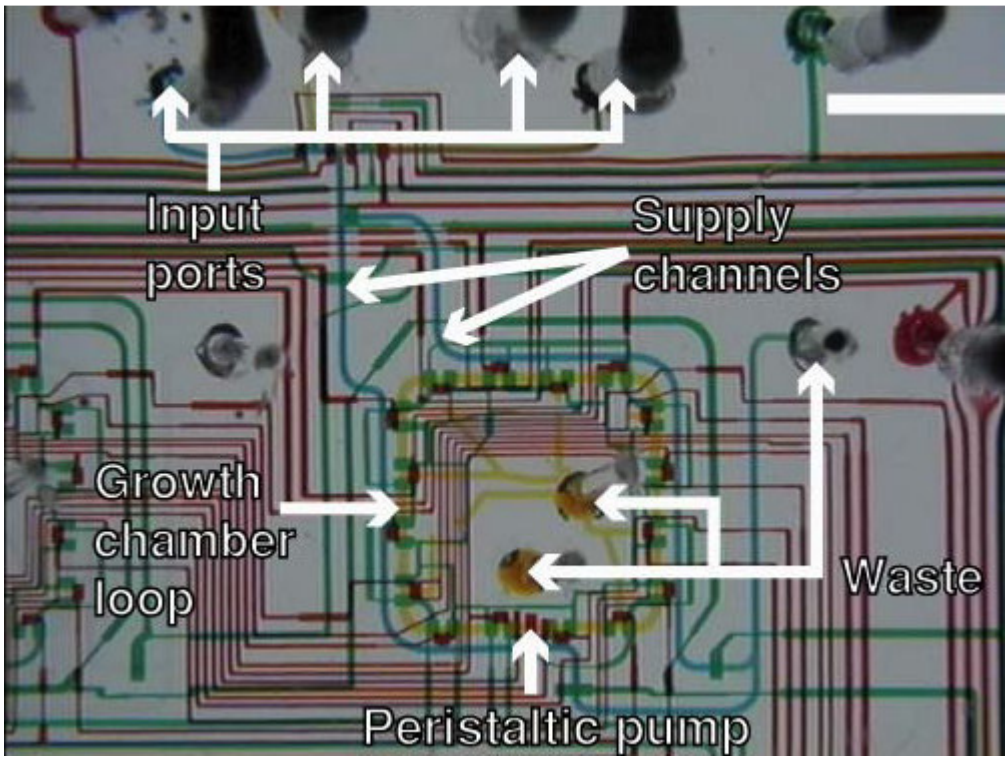


Figure 4-3 Microchemostat image with channels colored by food dye.

The growth chamber loop is colored with yellow food dye. The supply channels that bring fresh media to the reactor during cleaning events are colored in blue. The control lines for actuating the microfluidic valves throughout the chip are colored in red and green. Microchemostat image by Frederick Balagadde [83].

Each of these 16 segments of the growth chamber loop can be individually isolated from the rest of the reactor and the cells and media within the chamber can be evacuated through a waste outlet and replaced with fresh media without cells. In the microchemostat the emptying of one of these segments serves two purposes: (1) it is the mechanism of dilution of cells associated with establishing the continuous culture environment and (2) it

prevents the formation of biofilms by flowing fresh media though the segment to sheer cells from the wall or by flowing lysis buffer to actively kill any cells that have adhered to the walls. The microchemostat has two modes of operation (Figure 4-4): (1) "continuous circulation" of the cells and media in the reactor to maintain a well-mixed environment and (2) a "cleaning event" where one of the 16 reactor chambers is isolated, cleaned, and replaced with fresh media. During an experimental run of the microchemostat the reactor spends most of its time in continuous circulation mode punctuated by scheduled cleaning events that clean each of the 16 reactor chambers.

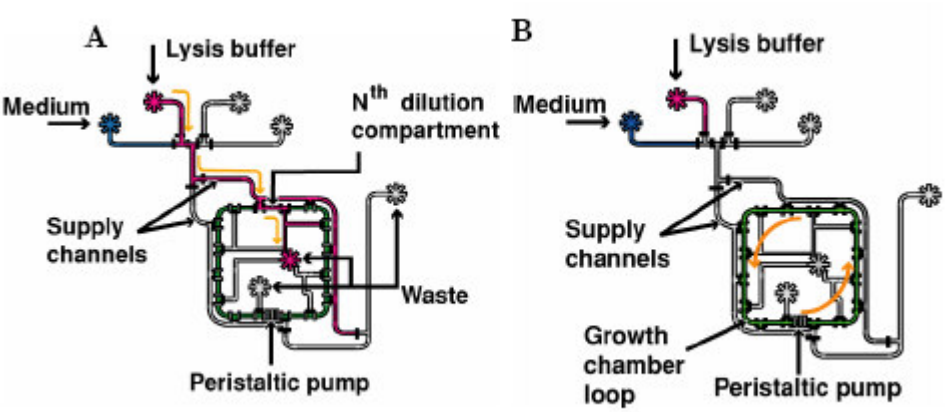


Figure 4-4 Two modes of operation of the microchemostat.

(A) During a "cleaning event" one of the 16 segments of the reactor is isolated and lysis buffer is flowed through the segment to eliminate any biofilm growth. The lysis buffer is represented in pink flowing from the inlet through the 'N^{th} dilution compartment' and out the waste outlet. Following the lysis buffer, fresh media will be flowed into the segment and then the microchemostat will move to (B) "continuous circulation" mode. In continuous circulation mode, the cells and media are mixed continuously around the growth chamber loop by the inline peristaltic pump. This mixing ensures that the fresh media added during cleaning events is mixed throughout the culture. Image by Frederick Balagadde [83]

4.3.2 Design and Fabrication of the Sortostat microfluidic chip

My goal in designing the Sortostat was to incorporate non-random dilution of cells so that unique selective pressures could be applied to cells growing in the reactor. To accomplish this I wanted to add a third mode of operation beyond the existing two in the microchemostat: a "sorting event" where a small fraction of the culture is isolated, imaged via automated microscopy, and either sent to the waste or released back into the reactor to continue growing. In theory, the sorting event could provide all the necessary dilution of the cell culture needed to maintain continuous culture growth and the cleaning event could be removed altogether thus dramatically simplifying the design of the device (for instance, by eliminating the need for the 16 individually addressable reactor segments). However, in practice the second function of the cleaning event, the prevention of biofilm formation, remains critical for long-term functioning of the Sortostat. Thus the Sortostat was designed to have three modes of operation: (1) continuous circulation, (2) cleaning event, and (3) sorting event.

In general I tried to avoid making significant changes to the design of the microchemostat when adding the new sorting event functionality, and went through several design variants (Appendix C) before settling on the final design (Figure 4-5). The final design removes the parallel operation of the reactors and reduces the number of reactors per chip from six to two. This design change was necessary to improve the reliability of the chip

and to address a possible issue with oxygen diffusion due to an overly dense network of control channels (Appendix C). Additionally, I added three new valves to each reactor to enable the isolation of an individual 100 x 100 µm sorting chamber with a volume $\approx 1/100^{th}$ of the total growth chamber loop volume (Figure 4-6). Due to the density of the control channels on the chip (dyed red and green in Figure 4-1), adding the additional three valves required a redesign of the layout of the control channels across the entire previously described chip [82]. The final dimensions of the Sortostat growth chamber loop were 10µm high, 100µm wide, and 11.5mm in circumference. However, the basic operation of the chip when undergoing continuous circulation or cleaning events remained identical to the microchemostat (Figure 4-4). I designed the Sortostat microfluidic chip in AutoCAD 2004 (Autodesk Inc) and the physical PDMS chips were fabricated by the Kavli Nanoscience Institute Microfluidic Foundry at the California Institute of Technology [93] (Methods, Appendix B).

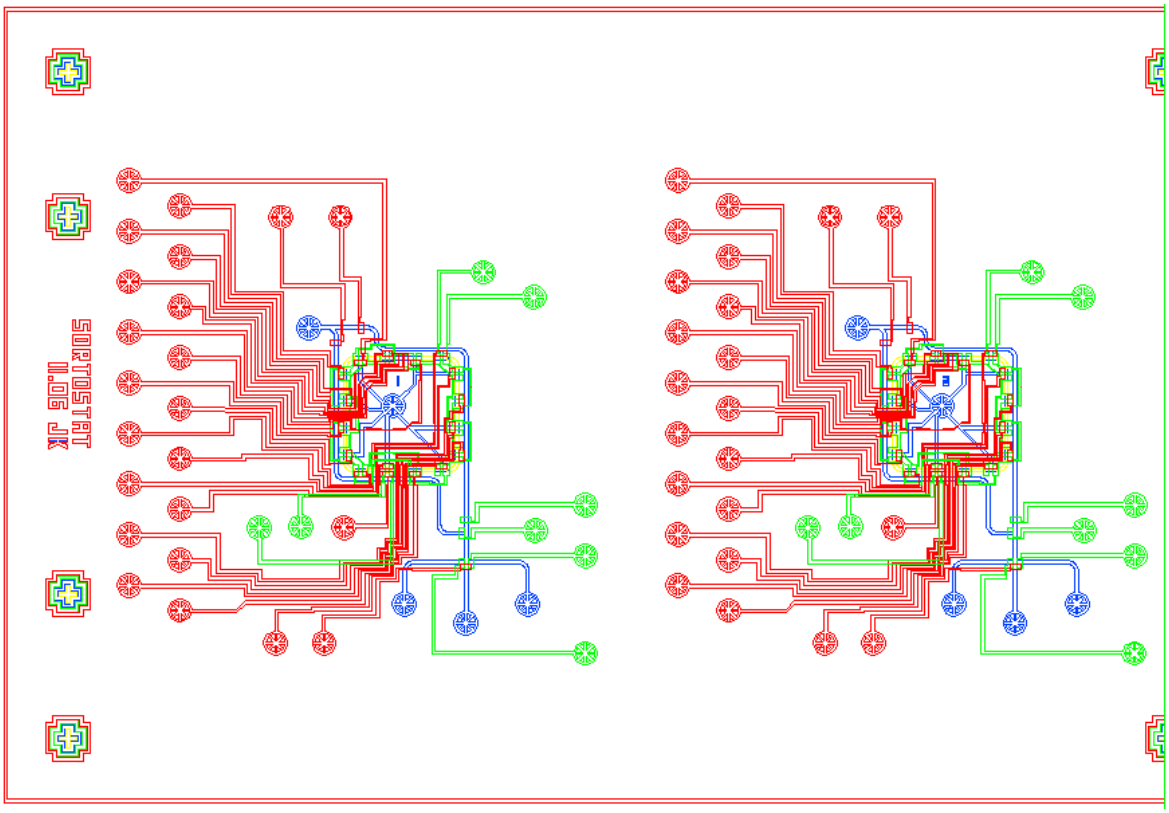


Figure 4-5 Final design for the Sortostat microfluidic chip.

Each chip contains two independent instances of the Sortostat. Thus push-down control lines are shown in red, the push-up control lines are shown in green, and the fluid lines are shown in blue. The larger green and red circles represent inlet ports where the micrfluidic control channels interface with macrofluidic, pressurized control lines containing water. In total 31 independent control lines are needed to operate the Sortostat. The blue circles represent ports where the microfluidic fluid lines interface with inlet lines for media or exit lines for waste.

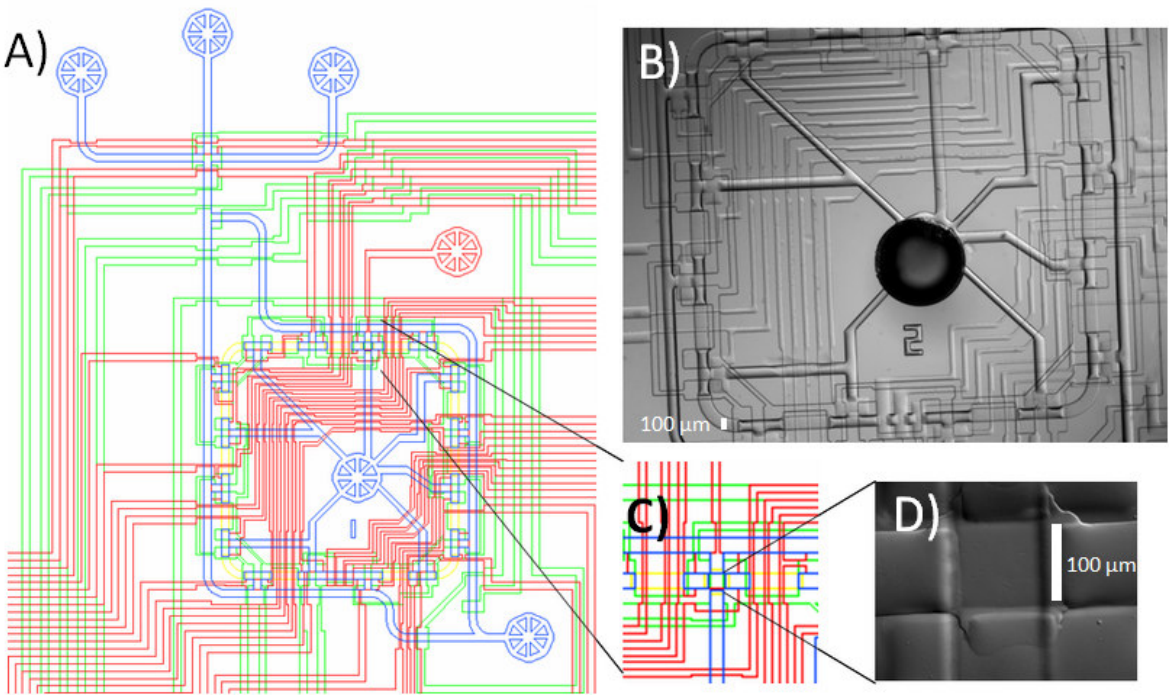


Figure 4-6 Addition of the sorting chamber to the microchemostat design

(A) Schematic of the entire Sortostat and (B) micrograph of the entire Sortostat. (C) Close-up of the sorting chamber (the small square in the center of the blue cross). There are 4 independently controlled valves surrounding the sorting chamber: two push-down valves (in red above and below) and two push-up valves (in green on the left and right). (D) A micrograph of the sorting chamber where the lower valve can be seen to be clearly sealed.

4.3.3 Sortostat automated platform.

The microfluidic chip is one piece of the Sortostat automated platform that is necessary for controlling the device, the platform includes: (1) a computer running custom LabView software [94] to control coordinate the overall operation of the Sortostat, (2) computer-controlled actuators for modulating pressure to the control lines, (3) a reservoir containing fresh media and a receptacle for collecting waste media and cells, (4) an automated microscopy platform for computer-controlled image capture, (5) a constant

temperature incubation chamber enclosing the chip, and (6) a custom MATLAB script for image processing and automated decision-making during sorting events (Figure 4-7). The Sortostat automated platform is almost completely automated, the only human intervention needed is the occasional (every 2 days) refocusing of the objective due to drift of the stage.

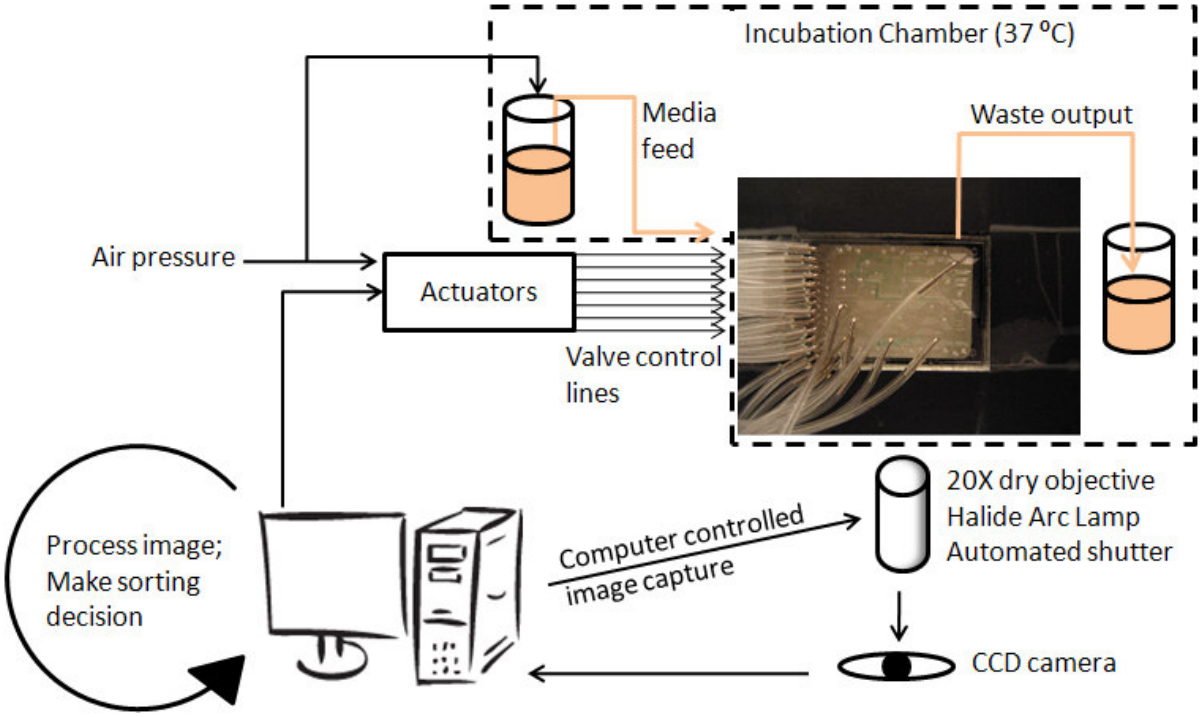


Figure 4-7 Schematic of the Sortostat automated platform.

The Sortostat automated platform is highly automated, allowing for multi-week experiments with little intervention.

4.3.4 Operating the Sortostat

Once the reactor had been inoculated with the cells, the user can specify a rate of cleaning events and sorting events to achieve a target overall dilution rate ($D_{overall}$) in the

reactor. The relationship between the rate of cleaning and sorting events and the overall dilution rate is given by the following equations:

$$D_{overall} = D_{clean} + D_{sort} \quad \text{(Eq. 4.1)}$$

The individual dilution rates are related to the volumetric flow rate (F) and the total reactor volume (V) by the follow equation:

$$D = F/V \quad \text{(Eq. 4.2)}$$

Since the flow of new media into the Sortostat occurs in discrete steps the volumetric flow rate (F) is defined by the size of the reactor segments used for cleaning ($\left(\frac{1}{16}\right)V$) and sorting ($\left(\frac{1}{100}\right)V$) combined with the rate of occurrence of cleaning events (δ_{clean}) and sorting events (δ_{sort}):

$$F_{clean} = \left(\frac{1}{16}\right)V * \delta_{clean} \quad \text{(Eq. 4.3)}$$

$$F_{sort} = \left(\frac{1}{100}\right)V * \delta_{sort} \quad \text{(Eq. 4.4)}$$

Equations 4.1-4.4 can be combined and simplified to provide the relationship between cleaning and sorting rates and the overall dilution rate in the Sortostat:

$$D_{overall} = \left(\frac{1}{16}\right) * \delta_{clean} + \left(\frac{1}{100}\right) * \delta_{sort} \quad \text{(Eq. 4.5)}$$

The overall dilution rate ($D_{overall}$) is a user-defined property, although the dilution rate needs to be less than the maximum growth rate of the cells or else the population will be washed out of the reactor [45]. In order to maximize the strength of the selective pressure that is applied to cells in the Sortostat, as much of the dilution as possible should be due to directed sorting events (D_{sort}) rather than cleaning events (D_{clean}) since cleaning events remove a random fraction of cells from the reactor. In all the experiments that follow I set the cleaning rate to a "minimum" value of 3.78 hr^{-1} (in other words, all 16 segments of the reactor are cleaned in about 4 hours 15 minutes). I found this cleaning rate to reliably prevent biofilm formation under the conditions used, while some slower cleaning rates led to occasional wall growth. In all the experiments described here I used a dilution rate of 1 hr^{-1} which leads to a sorting rate of 37.7 hr^{-1} (one sorting event every 1.6 minutes).

Based on the user-specified cleaning rate and sorting rate, the Sortostat automated platform automatically initiates properly timed cleaning and sorting events. A cleaning event is straightforward: one of the 16 reactor segments is isolated by closing and opening the appropriate valves and fresh media is flowed through the segment flushing the cells out the waste. The LabView software tracks the number of cleaning events and cleans the appropriate segment, working around the entire reactor approximately every 4 hours and 15 minutes in the experiments described here.

A sorting event is more complicated than a cleaning event and involves two steps: (1) the sorting chamber is isolated by closing the appropriate valves and images are captured and stored using the automated microscopy platform and (2) the images are processed in real

time by a MATLAB script and a decision is made by the LabView software to flush the cells in the sorting chamber out the waste. If instead a decision is made to return the cells in the sorting chamber back into the larger population of cells then I refer to this as a "screening event" rather than a sorting event as there is no dilution of the culture. However, in order to ensure a constant dilution rate it is important that sorting events happen with a regular frequency independent of what cells are randomly captured in the sorting chamber. For instance, in the experiments described here I decided that for every two screening or sorting events one of the two must be a sorting event. Therefore, if the first of two events is a screening event (in other words, if the computer decides to release the cells rather than flush them out the waste) then the second must be a sorting event by definition. Even if the computer would normally choose to release the cells based on the image processing, the algorithm forces the cells to be sent to the waste in order to ensure a constant dilution rate. The opposite case is also true: if the first in a pair of events is a sorting event then the second event will be a screening event by definition in order to maintain a constant dilution rate.

4.3.5 Models of Sortostat function

Models for normal chemostat operation have been well described previously [45, 84, 85]. In particular, Feredrick Balagadde extended the standard logistic model for growth in a chemostat [45] to account for the discrete dilution events in the microchemostat [83]. These previous analyses largely focus on the growth of cells following inoculation until they achieve steady-state growth in the chemostat as well as on the expected steady-state cell

density that will be reached at steady state. While inoculation followed by growth until steady-state occurs in the Sortostat, the new features of the device (directed sorting) are not turned on until steady-state growth has been achieved. Thus, in the models associated with the experiments that follow (see also Chapter 5) I will focus my analysis on the regime where the cells have achieved steady-state growth rate in the reactor and are at a constant density.

4.3.6 Proof of principle experiments

In order to demonstrate the functionality of the Sortostat I ran a number of experiments including: (1) "normal" chemostat operation without applying any selective pressure, (2) applying a time-varying selective pressure to a subpopulation of cells in the population, and (3) selecting for the maintenance of a slower growing subpopulation that would otherwise be washed out of the reactor. In each of the experimental runs I inoculated the Sortostat with a co-culture of two strains of MC4100 *E. coli*: one containing a constitutive CFP generating device (BBa_I13602) and one containing a constitutive YFP generating device (BBa_I6032). Both devices were carried on a high-copy plasmid, pSB1A2, and the growth rates of the two strains were measured to be identical within experimental error (Appendix C). The number of CFP and YFP expressing cells in the sorting chamber were counted by automated microscopy and image processing with a custom MATLAB script (Appendix C).

4.3.7 Demonstration of chemostat operation and verification of expected statistical distributions

I first wanted to demonstrate that the Sortostat was capable of "normal" chemostat operation by growing cells in the reactor without applying a directed selection. The LabView software in this case simply made a random decision about whether to send cells in the sorting chamber to the waste. Under these conditions the Sortostat performs similar to the microchemostat (Figure 4-8 and 4-9).

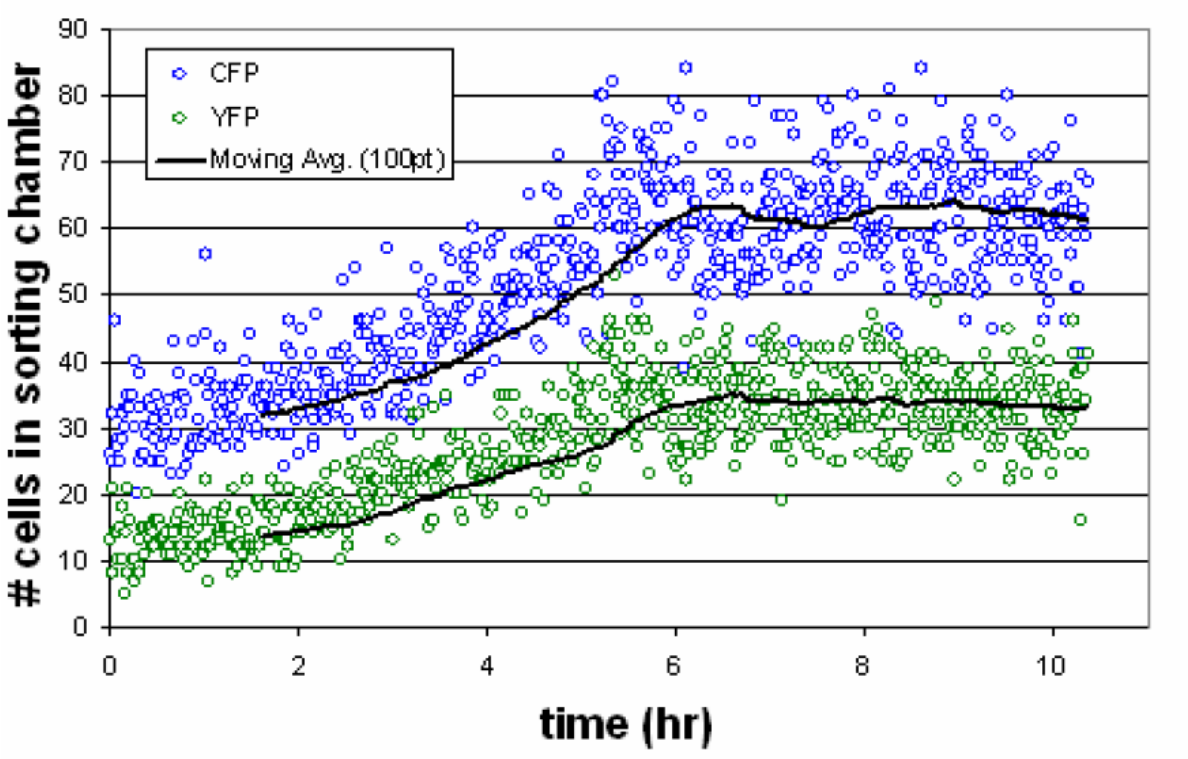


Figure 4-8 Demonstration of operating the Sortostat as a normal chemostat

Cells expressing CFP or YFP were co-cultured in the Sortostat and grown with a constant dilution rate but random sorting in order to demonstrate that the Sortostat can operate like a"normal" chemostat. For each time point two images were captured of the sorting chamber

and processed by automated image processing algorithms to count cells: one image was captured with a CFP emission filter to count CFP cells (blue circles) and the other was captured with a YFP emission filter to count YFP cells (green circles). The cells reach steady state growth by six hours as indicated by a leveling off of the cell counts.

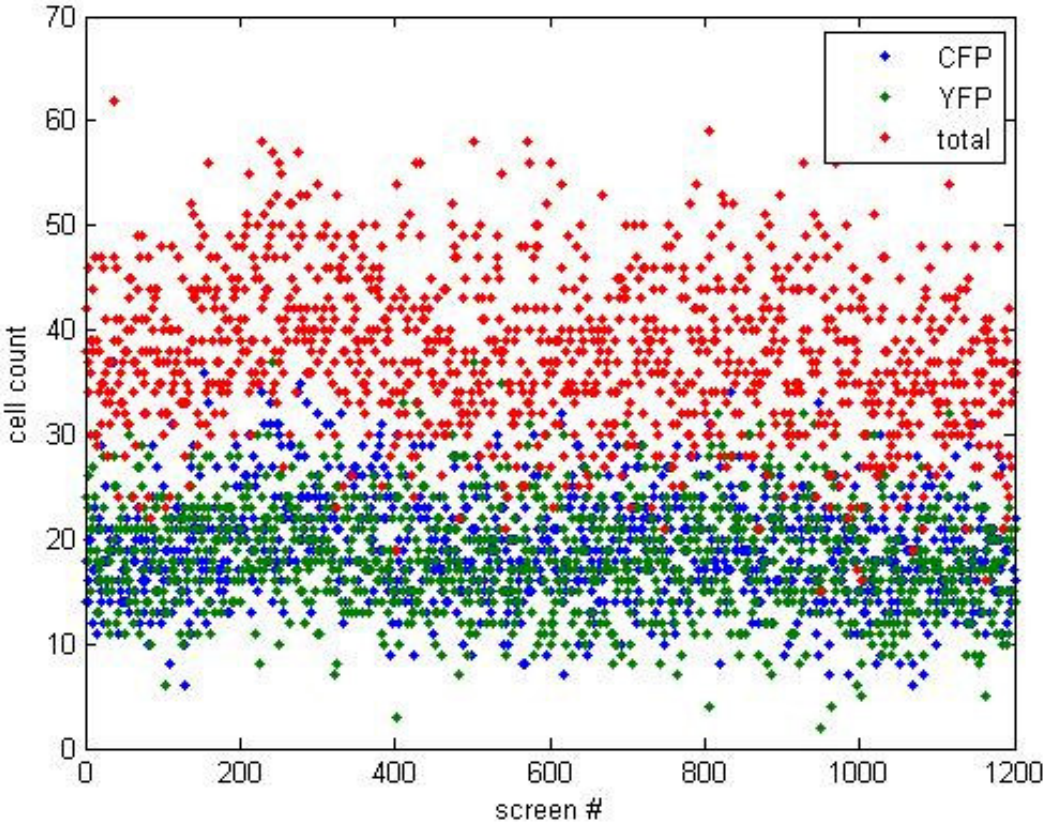


Figure 4-9 Demonstration of operating the Sortostat as a normal chemostat

A segment of a longer run where sorting is allowed to proceed at random. The earlier time points including the initial rise to steady state are not shown in the plot. The data here represents a 20 hour time course. For each time point two images were captured of the sorting chamber and processed by automated image processing algorithms to count cells: one image was captured with a CFP emission filter to count CFP cells (blue dots) and the other was captured with a YFP emission filter to count YFP cells (green dots). These two counts were summed to provide a total count in the sorting chamber (red dots). The total cell count remains relatively constant over the entire segment demonstrating that the cells are in normal, steady-state chemostat growth.

In order to better understand the performance of the Sortostat, I verified that experimental results agreed with expected theoretical statistical distributions for the cells captured in the sorting chamber. I used the data shown in figure 4-9 to conduct this analysis. First, I evaluated that the distribution of the total number of cells that were captured in the sorting chamber. Assuming that the growth chamber loop is well-mixed then the number of cells captured in the sorting chamber can be modeled as a binomial distribution where the probability of capturing k cells in the sorting chamber is given by the binomial equation:

$$P(K = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{(Eq. 4.6)}$$

Where n is the total number of cells in the reactor (3800 on average in this experiment) and p is the size of the sorting chamber relative to the total volume of the growth chamber loop (1/100). I can then solve for the probability mass function and compare it the histogram of the data (Figure 4-10). There is good agreement between the predicted distribution and the experimental data.

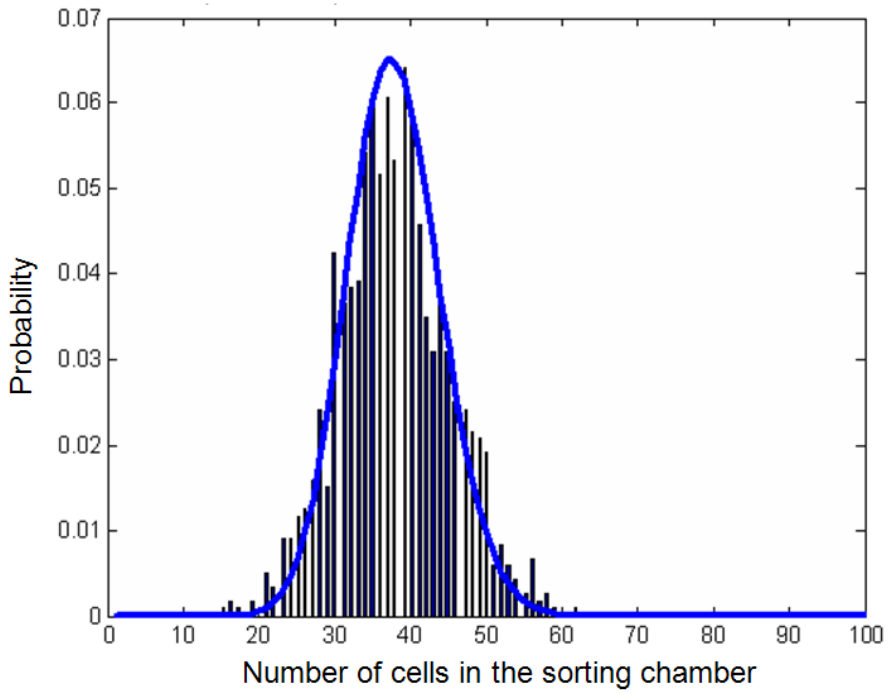


Figure 4-10 Comparison of model distribution to experimental results for the number of cells captured in the sorting chamber.

There is good agreement between the predicted binomial distribution (blue curve) for the number of cells captured in the sorting chamber and the histogram of experimental data (blue bars).

I can extend this analysis to also evaluate the expected number of cells expressing CFP captured in the sorting chamber based on the average fraction of cells expressing CFP in the reactor. Since the number of cells expressing CFP in the sorting chamber (c) will be correlated with the total number of cells in the sorting chamber (k) I will need to solve for the joint probability distribution:

$$P(K = k \; and \; C = c) = P(C = c|K = k)P(K = k) \quad \text{(Eq. 4.7)}$$

I previously solved for $P(K = k)$ in Eq. 4.6, so I only need to solve $P(C = c|K = k)$ that can be modeled as a binomial distribution:

$$P(C = c|K = k) = \binom{k}{c} p^c (1-p)^{k-c} \quad \text{(Eq. 4.8)}$$

Where p is the average percentage of cells expressing CFP in reactor (50%). I can then solve this distribution and compare the predicted joint distribution against the experimental data to find a good agreement (Figure 4-11).

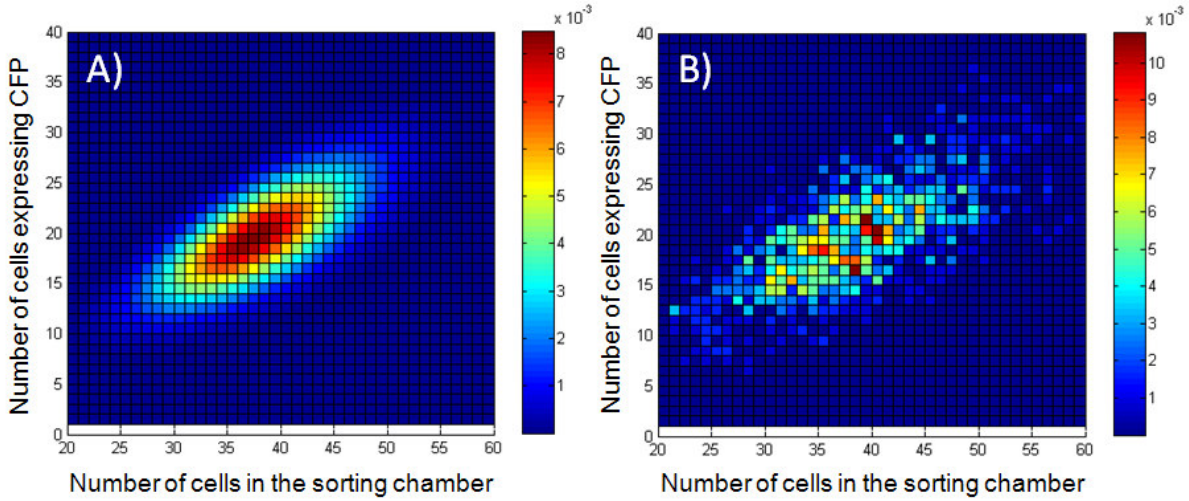


Figure 4-11 Theoretical joint probability distribution of the total number of cells in the sorting chamber and the number of CFP expressing cells in the sorting chamber agrees with experimental data.

Comparison of (A) theoretical and (B) experimental joint probability distributions, the color bar represents the probability of capturing a particular pair of the total number of cells in the sorting chamber and the number of CFP expressing cells in the sorting chamber. There is good agreement between the predicted distribution and the experimental data.

Accurate theoretical distributions of the number of cells expressing CFP or YFP captured in the sorting chamber will support the development of more complicated models of

the overall function of the Sortostat (Chapter 5). In particular, the strength of the selective pressure that can be applied to a subpopulation (such as CFP expressing cells) will be dependent on the distribution of cells from that subpopulation that are captured in the sorting chamber. For example, a wider distribution of the number of CFP expressing cells in the sorting chamber would allow for a higher selective pressure to be applied to the CFP expressing cells. To better understand the relationship between the distribution of CFP cells and the selective pressure consider the extreme case where there is no variation in the number of CFP cells captured (in other words, an infinitely narrow distribution). In that scenario I would be unable to apply any selective pressure to the CFP subpopulation as every sorting event would remove the same number of CFP cells.

4.3.8 Demonstration of time-varying selective pressure in the Sortostat

In order to demonstrate that the Sortostat is capable of complicated screens I again co-cultured cells expressing CFP or YFP in the Sortostat. I applied a time-varying selective pressure to a subpopulation of the cells: switching from selection against the CFP expressing subpopulation to selecting in favor of the CFP expressing subpopulation. The specific sorting approach used here was: (1) the sorting chamber was isolated by closing the appropriate valves, (2) two fluorescent images were captured of the cells in the sorting chamber: first with a CFP filter and then with a YFP filter, (3) the images were processed by a MATLAB script to count the number of CFP and YFP cells, (4) the fraction of CFP cells was compared to the running average of the previous 15 CFP fractions measured, (5) if the

platform is set to be sorting against CFP then if the fraction of CFP cells is above the running average the cells are flushed out the waste; if the platform is set to be sorting in favor of CFP then if the fraction of CFP cells is above the running average the cells are released back into the growth chamber loop.

I set the Sortostat automation platform to sort in favor of CFP cells until the fraction of CFP expressing cells in the reactor reached 70%. At 70% CFP expressing cells the Sortostat would flip the direction of selection to sort against CFP expressing cells. Once the percentage of CFP expressing cells in the population had been driven down to 20% the direction of selection would be flipped again to now sort in favor of the CFP expressing cells. The selective pressure was flipped repeatedly between hours 30 and 150 of a 180 hour experiment resulting in a periodic rise and fall in the percentage of CFP expressing cells in the reactor (Figure 4-12). Furthermore, the total number of cells (red, Figure 4-12A) remains relatively constant throughout the 180 hour experiment (after an initial equilibration phase of 30 hours, Appendix B) suggesting that the sorting approach used here was able to maintain a constant overall dilution rate. Finally, a 20 hour region (150 to 170 hours) was included where no directional selection was applied to the population, the number of CFP expressing cells in the population remains relatively constant during this period as expected.

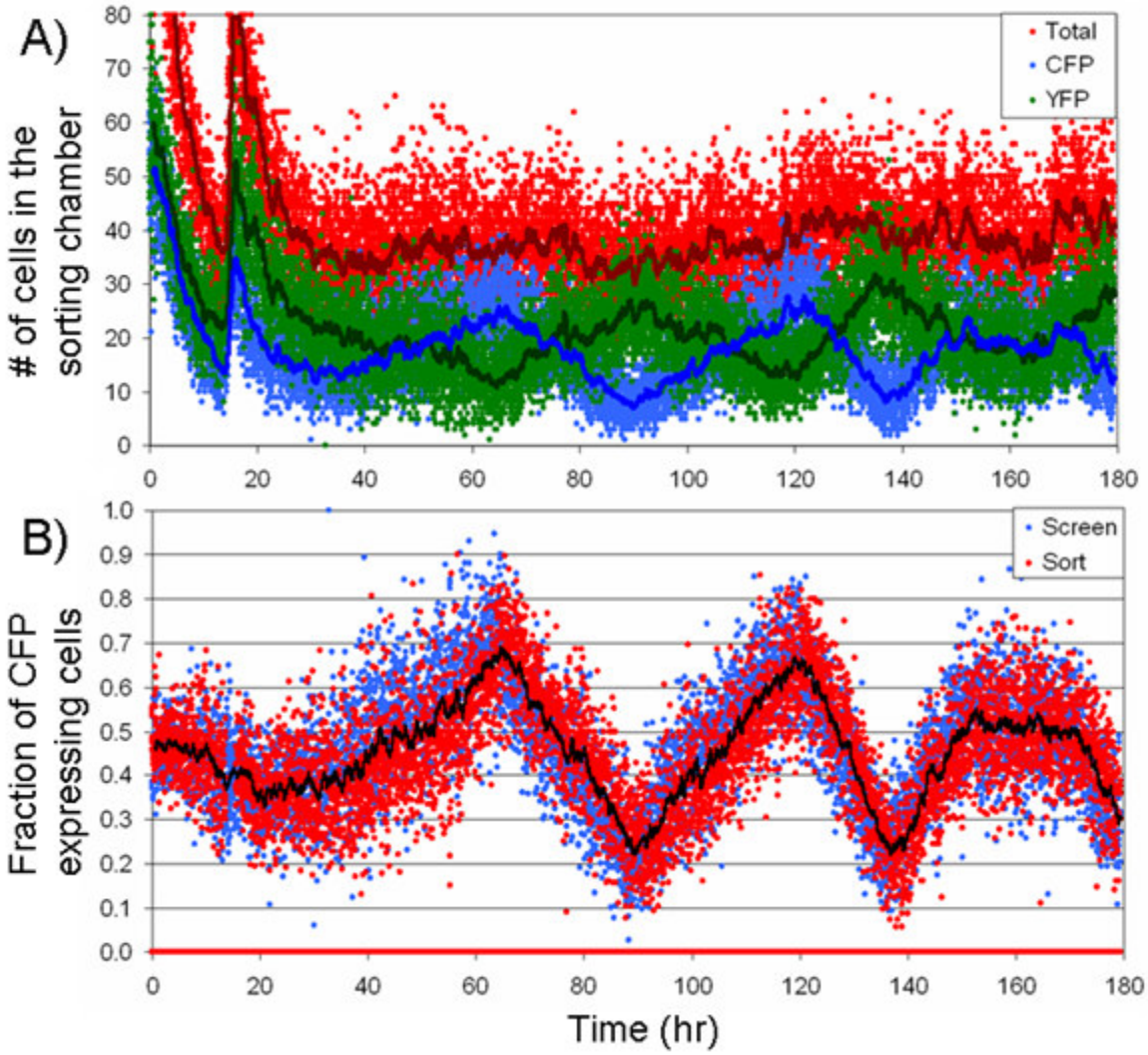


Figure 4-12 Demonstration of time-varying selection using the Sortostat

A 180 hr run of the Sortostat demonstrates several different modes of operation of the device: an initial equilibration phase where the cells are approaching steady state (<30 hours), a region of time-varying selection (30 to 150 hours), and a region with no selection (150 to 170 hours). (A) For each time point there are three values shown: a blue dot represents the number of cells expressing CFP in the sorting chamber, a green dot represents the number of cells expressing YFP, and a red dot represents the total number of cells (CFP+YFP). The total number of cells remains fairly constant after the initial equilibration phase as expected for a constant overall dilution rate. Darker lines are the 15 point moving averages for each data type. (B) For each time point there is a single value shown: the percentage of CFP expressing cells in the sorting chamber. The color of the points represents

whether that particular sorting chamber was sent to the waste (Sorting event, red) or released back into the growth chamber loop (Screening event, blue). The black line represents the 15 point moving average. In the first 30 hours, the cells are approaching steady state and go through one spike in the number of cells due to a manual intervention that turned off the sorting events at approximately hour 18. The remainder of the run proceeded normally. In hours 30 to 150 the number of CFP and YFP cells can be seen to periodically rise and fall (A). Note that the majority of sorting events (B, red points) fall above or below the average (black line) depending on the direction of sorting as expected. In hours 150 to 170 directional sorting was turned off and as a results the number of CFP and YFP cells remain constant (A) and the sorting events (B, red points) are evenly distributed around the average (black line).

4.3.9 Rescue of a slower growing subpopulation in the Sortostat

To demonstrate the application of a unique selective pressure in the Sortostat I applied a selective pressure in favor of a slower growing subpopulation of cells in the reactor and were able to promoter this subpopulation for 150 hours. Although the growth rate of the CFP and YFP expressing cells used in the experiments here were measured to be identical (Appendix B), on occasion I found that the one or the other cell type grew at a faster rate when the cells were co-cultured in the Sortostat (possibly due to faster growing mutants). In this experiment it was found that the CFP expressing cells were at a disadvantage to their YFP expressing counterparts. Over the first 100 hours of the experimental run the fraction of CFP expressing cells in the reactor dropped steadily. The region was punctuated by rises in the fraction of CFP expressing cells as I was applying a time-varying selection at points during this period, however the overall trend clearly shows that the CFP expressing cells were being out-competed by the faster growing YFP expressing cells (Figure 4-13). At 100 hours I changed the direction of the selective pressure to favor CFP expressing cells for the

duration of the experiment. Over the next 50 hours the CFP expressing cells were driven from 20% of the total cell population to over 80% of the population.

The capacity to maintain slower growing subpopulations is a valuable tool for both screening and characterization of engineered biological systems. Complex synthetic biological systems often result in a reduction in the growth rate of the cellular chassis that houses the system and as a result these systems are susceptible to failure when fast-growing, dysfunctional mutants arise and outcompete functioning systems [49]. The Sortostat could provide a platform for maintenance of functional systems over long periods of time so that these systems could be better characterized and improved. Furthermore, when selecting from libraries of biological systems, mutants with desired phenotypes may be slower growing and the Sortostat will provide a mechanism to enrich the culture for these mutants.

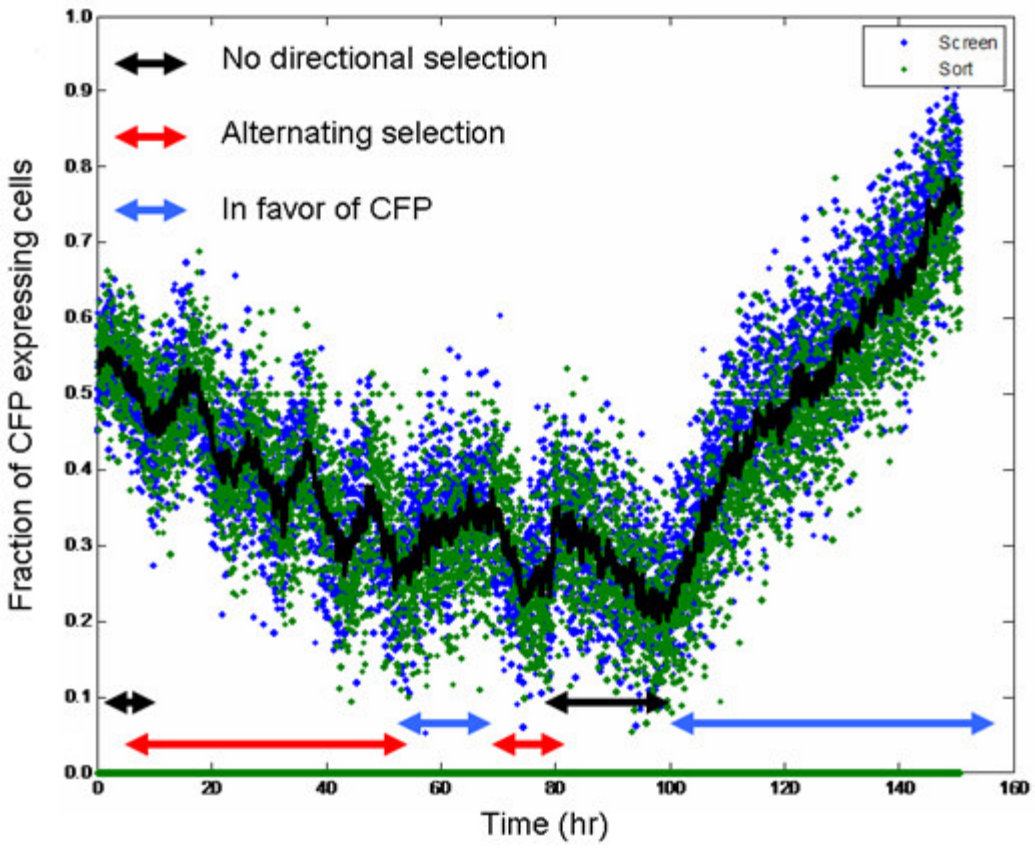


Figure 4-13 Demonstration of the rescue of a slower growing subpopulation of cells using the Sortostat.

Each point on the plot represents the percentage of CFP expressing cells in the sorting chamber of a screening (blue) or sorting (green) event during a 150 hour Sortostat experimental run. In the first 100 hours the Sortostat was run under no selection (black arrows), alternating selection approximately every five hours (red arrows), and selection in favor of CFP (blue arrows). During the periods with no selection (black arrows) it is clear that the fraction of CFP expressing cells in the reactor fell due to a slower growth rate than their YFP expressing counterparts. For the final 50 hours of the run I set the Sortostat to select in favor of CFP expressing cells and demonstrated that I could rescue this slower growing population, driving the percentage of cells expressing CFP from 20% to 80% of the population.

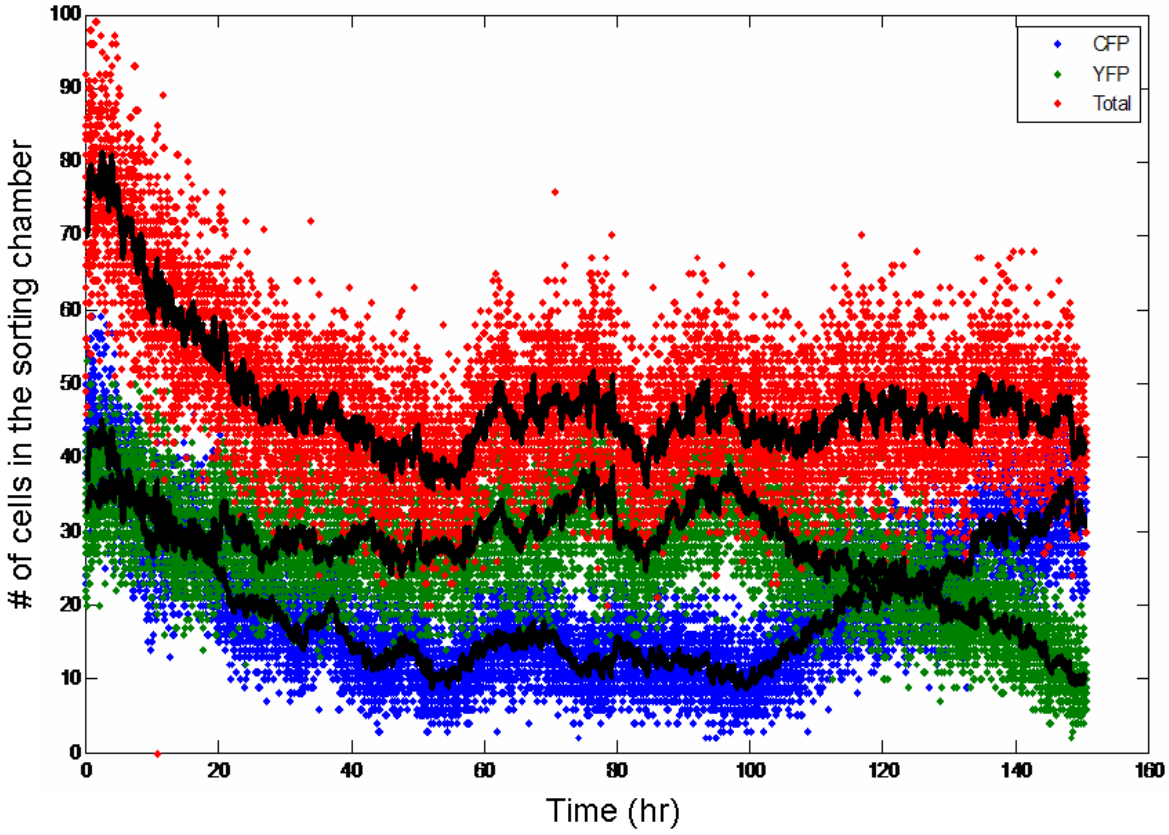


Figure 4-14 Demonstration of the rescue of a slower growing subpopulation of cells using the Sortostat.

For each time point there are three values shown: a blue dot represents the number of cells expressing CFP in the sorting chamber, a green dot represents the number of cells expressing YFP, and a red dot represents the total number of cells (CFP+YFP). The slower growing CFP population is promoted starting at 100 hrs until it reaches nearly 80% of the total cell population.

4.3.10 Improvements for future versions of the Sortostat

The Sortostat is a first generation device integrating cell sorting into a continuous culture bioreactor. There are many opportunities for improving the device, including: (1)

improved methods for countering wall growth would allow for a simpler reactor design, (2) more sophisticated sorting approaches that would allow for stronger selective pressure, and (3) shallower channels that would allow for better imaging for more complicated selections.

The current method for countering wall growth relies on independently addressing 16 segments of the growth chamber loop. Addressing these segments independently requires the majority of the control lines used to operate the Sortostat. If a passive approach to preventing biofilm growth in a PDMS microenvironment were developed (for instance, a more effective non-adhesive wall coating) the Sortostat could be operated with roughly 10 control lines rather than 31. Lowering the number of control lines needed would speed setup time, reduce opportunities for fabrication error, and simplify device operation. Finally, I expect that a simplified design would help to address the most common failure modes of the device (Appendix C).

The current sorting approach groups sorting and screening events into pairs. If the first event of the pair is a sorting event then the second is required to be a screening event, and vice versa. Ensuring that sorting events happen with a reliable average rate of occurrence is necessary to ensure an approximately constant dilution rate and keep the cell density stable. However, screening and sorting events could be grouped into larger sets to enable more chambers to be evaluated before forcing a sorting event. In general, if more screening events occur between each sorting event than the selective pressure applied will be stronger. There is a practical limit based on the time it takes to capture and process images;

however with the current Sortostat automated platform, a screening event takes approximately 15 seconds allowing for multiple screening events for each sorting event. Additionally, the Sortsotat could be run in a turbidostat mode [95] in order to enable dilution rates that are closer to the maximum cellular growth rate. Higher dilution rates would reduce the cell density in the reactor and enable stronger selective pressure to be applied to the population. Finally, the Sortostat could be run as a completely new type of reactor without stable dilution or constant cell density. This might allow for faster selection, though the constant environmental conditions of the Sortostat would be lost.

The Sortostat enables selection based on any information that can be gleaned from processing an image of cells captured in the sorting chamber. Thus, my ability to apply complex selective pressures (for instance, pressures based on cell morphology) will be a function of the quality of the images that can be captured by the Sortostat automated platform. The current depth of the channels in the Sortostat (10μm) exceeds the focal depth of typical objectives needed to view *E.coli* cells. Work by others has already begun to implement shallower channels for improved imaging (Appendix C); these channels could be implemented in the next version of the Sortostat.

4.4 Conclusions

I described a new microfluidic device, the Sortostat, that integrates a cell sorting chamber with a previously published microscope-mounted microfluidic chemostat. I described the design and construction of the Sortostat microfluidic chip as well as the

automated platform used to operate the chip. I validated that the number of cells in a subpopulation (CFP expressing cells) captured in the sorting chamber during an experimental run matches the expected theoretical distributions based on assumptions of a well-mixed reactor environment. Knowledge of these distributions lays the foundation for future modeling work to better describe the performance limits of the Sortostat. I demonstrated that the Sortostat is capable of applying unique selective pressures by applying a time-varying selection to modulate the concentrations of two competing subpopulations of *E.coli* engineered to express either CFP or YFP. Furthermore, I demonstrated that the Sortostat can be used to select in favor of slower-growing subpopulations by stably maintaining a disadvantaged bacteria in the Sortostat for many hours – an application that may be valuable for the characterization of engineered biological systems that are prone to loss of function due to faster-growing mutants [49].

Chapter 5. Future Work

5.1 Introduction

In the work described in this thesis I have begun to address some of the bottlenecks in applying directed evolution to the construction of functional devices (Chapter 1). However, many challenges remain in each of the areas outlined.

I was able to use the PoPS input/output characterization and screening plasmid (pSB1A10) to repair the function of a dysfunctional inverter (BBa_Q04400), however repeated attempts to repair other inverters were unsuccessful. Anecdotally, I expect that inverter libraries could be generated with greater fractions of functional variants if the libraries were constructed by assembling parts (for instance, inserting a library of RBSs) rather than by building libraries by varying the sequence across the entire device. However, the experiments described in this thesis are insufficient to determine if a parts-based assembly approach to building libraries is superior to randomly varying the sequence across the entire device. Future work to explore this question would be valuable for researchers engineering devices. To begin to evaluate this question, I outline a plasmid-based scaffold that will allow for tuning of inverters via insertion of part libraries (Section 5.2).

The measurement kits and reference standards described in Chapter 3 provide a mechanism to make reliable measurements of promoters and RBSs in standard units, however the kits point to the need for improvements in the specifications of the parts

themselves in order to allow for more reliable measurement (for instance to standardize the transcriptional start site and downstream DNA sequence). Work could begin immediately to establish a part specification for promoters (Section 5.3) and other part types. These specifications are necessary to ensure that part function is as independent of context as possible as well as to allow for reliable characterization using the measurement kits. In general, biological engineers should begin to make engineering decisions that might limit the options of part designers but allow for greater reliability in part function. For example, in the "fauxmoter" promoter specification described below the sequence of the +1 to +20 region of the promoter is fixed and as a result an mRNA designer would be unable to alter the first 20 bp on the 5' end of any mRNA part that was expressed by a promoter that met this specification. While this clearly limits the options for mRNA design, it will likely allow for promoters that operate more reliably independent of their context. Such a trade-off should be evaluated by constructing and testing promoters that meet the specification and then distributing them to the community to evaluate their efficacy.

I Chapter 4 I described a number of initial proof of principle experiments to demonstrate the operation of the Sortostat, however I have not yet attempted to use the Sortostat to screen a library of systems to isolate functional variants. In future experiments the Sortostat could be used to select among libraries of systems with more complicated functions such as oscillators with time-varying outputs. Selection for a functioning oscillator with a particular period would require that the Sortostat is capable of applying sufficient selective pressure at the time scale of the desired oscillation frequency. In order to better

understand the performance limits of the device and establish if such a selection if feasible, I have begun work on a stochastic simulation and model of Sortostat operation (Section 5.4).

5.2 Scaffold for tuning inverters by inserting libraries of parts.

[The work discussed here is being carried out in collaboration with Felix Moser]

In Chapter 2, I described finding that the previously tuned tetR-based inverter (BBa_Q04401) did not function on the newly constructed low copy characterization and screening plasmid (pSB3K10). I could have again attempted to construct a library of inverters via mutagenic PCR and screened for a variant that functioned across the range of PoPS inputs provided by pSB3K10. However, based on my previous experience tuning BBa_Q04401 it seemed likely that I could construct a library containing more functional variants by varying the RBS directly rather than mutating the entire inverter. Furthermore, in the future I may want to vary the promoter part in the inverter as well in order to modulate the output PoPS level of the inverter. To address these goals, I designed a plasmid-based scaffold (Figure 5-1) for tuning the RBS and promoter of tetR -based inverters. This scaffold can be used to test whether libraries constructed via combinatorial part assembly might be superior to libraries generated at the level of the DNA sequence. It may also be practically useful for tuning inverters to meet a future standard in support of functional composition [13]. I have also designed similar scaffolds for lacI and cI-based inverters (not shown).

Felix Moser has constructed the tetR-based scaffold and is in the process of inserting a set of test RBSs and the BBa_R0040 promoter into the scaffold as a control to verify that the scaffold is functioning properly. Once the basic function of the scaffold has been confirmed, larger libraries of promoters or RBSs could be inserted into the scaffold in order to screen for inverters with desired functionalities.

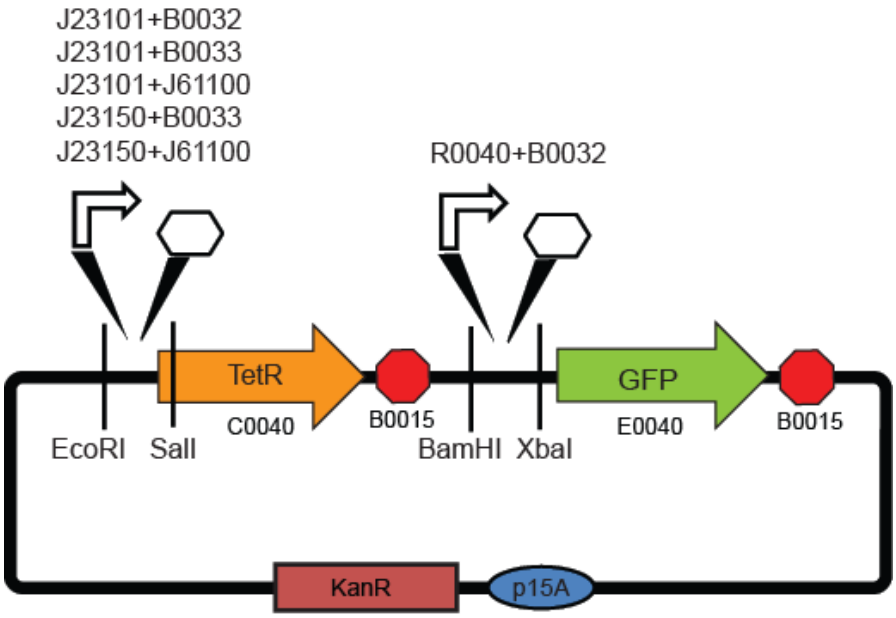


Figure 5-1 Schematic of the scaffold for tuning tetR-based inverters.

The scaffold contains two pairs of restriction sites for inserting promoter-RBS pairs. The first pair of restriction sites (EcoRI / SalI) is upstream of the tetR coding region and is designed to receive a constitutive promoter followed by an RBS library for tuning the RBS part within the inverter. The constitutive promoter will provide either the "Low" or "High" input signal to the inverter depending on the promoter chosen. Constitutive promoters were chosen rather than an inducible promoter in order to improve the reliability of the device. The second pair of restriction sites (BamHI / XbaI) is downstream of the B0015 terminator in the inverter and is designed to accept a library of promoters regulated by the Tet repressor and a standard RBS for the downstream GFP expression device. This site is used for tuning the promoter associated with the inverter. The RBS was included in the insert rather than on

the scaffold to provide flexibility for future reporter standards or in case future inverters were designed to input and output transcriptional signals rather than translational signals. The restriction sites are designed so that the inserted promoter-RBS pairs form the appropriate BioBrick junctions between parts.

5.3 Part specification for BioBrick promoters

[The work discussed here is being carried out in collaboration with Joey Davis]

Researchers frequently extract promoters from their natural context and use them to drive the production of exogenous, non-native transcripts. Promoter design is often done in an ad-hoc fashion using heuristics that vary from lab to lab. For example, a common heuristic for bacterial promoter specification is to include 200 nucleotides upstream of the open reading frame (ORF) start site. Typically, the activity of the promoter is measured based on the expression of downstream product and if the promoter is moved to a new context the activity of the promoter may change thus invalidating previous characterization. To improve the reusability of promoters an improved part specification for promoters is needed. Here I will outline a draft specification for BioBrick promoters that meets the following criteria:

1) The promoter should exhibit consistent activity independent of context. In other words, the number of polymerases escaping the promoter should remain constant independent of the nucleotide sequence immediately adjacent to the promoter.

2) The promoter should have a defined transcriptional start site that is identical for all BioBrick promoters.

3) The specification should allow for the generation of constitutive and regulated promoters spanning multiple decades in strength

4) Promoters should conform to an idempotent cloning strategy

In particular, the first two criteria for context-independent activity and a well-defined transcription start site have not been met by the current BioBrick promoter collection. To meet the four criteria outlined above, Joey Davis and I have outlined the following proposed specification for BioBrick promoters:

1) The promoter part spans from -100 to +20 (120 nt in length) at a minimum, and is flanked by the standard BioBrick restriction enzyme sites.

2) The -35 and -10 regions interact with a σ70 type sigma subunit.

3) The region between -10 and +1 is of the form: TATTATnnnnBCAT

-10 +1

The -100 and +20 positions were chosen as the minimum size for the promoter in order to insulate the promoter activity from the effect of adjacent sequence. Specifically, the -100 position was chosen because the bulk of known transcription regulatory elements in *E.coli* bind within this region [96]. The +20 position was chosen since it is known that the sequence in the +1 to +20 region effects the rate of promoter escape and thus the promoter activity [97]. We decided to limit the specification to σ70 promoters in order to simplify this initial specification, however we expect it could be expanded later for other promoter types.

Finally, we defined the -10 to +1 region in order to standardize the transcriptional start site for all BioBrick promoters.

5.3.1 "Fauxmoters"

The promoter measurement kit and reference standard described in Chapter 3 rely on promoters with identical transcriptional start sites and identical sequences downstream of the transcriptional start sites in order to ensure that the mRNA expressed by each of the promoters tested is identical to the mRNA produced by the reference standard promoter. Identical mRNA allows us to cancel the mRNA degradation rate and the translation rate of immature GFP from mRNA in order to simplify our model. This simplification is critical because it allows measurement in comparable units (SPUs) without needing to directly measure mRNA levels. To ensure that all promoters have a fixed transcriptional start position and fixed sequence downstream of the transcriptional start site a more stringent standard promoter definition is needed than the one described above.

These promoters will meet all the criteria defined in the BioBrick promoter standard described above, however they will also have a defined +1 to +20 region that is identical for all such promoters. Per the suggestion of Tom Knight, we will refer to these as "fauxmoters". All fauxmoters could then be measured using the measurement kits described in Chapter 3 or via future kits with alternate protein-based reporters of gene expression. However, if the fauxmoter specification is not adhered to then researchers measuring promoter activity will

need to quantify the rate of mRNA degradation for the different mRNAs produced by each non-standard promoter.

5.4 Models for describing the operation of the Sortostat

Mathematical models describing the operation of the Sortostat would be valuable for defining the performance limits of the device, as well as for guiding researchers in deciding the appropriate device settings to achieve a target selective pressure. I have done some early modeling work, however the models need to be improved and further explored to better understand the Sortostat performance limits.

In brief, I built a simple stochastic model that samples from the distributions described previously (Chapter 4) in order to simulate an experimental run of the Sortostat. This model assumes that the total number of cells in the Sortostat remains constant, and does not directly simulate cell growth in the reactor during continuous circulation mode. Instead, cell growth is only simulated following a dilution event (sorting event or cleaning event). In this rudimentary model for cell growth, after a sorting or cleaning event the cells that were removed from the reactor are instantaneously re-grown based on the updated ratio (in other words, the ratio calculated after the loss of the cells from the cleaning or sorting event) of CFP expressing cells in the entire reactor.

When the model simulates a sorting/screening event, the total number of cells captured in the sorting chamber and the fraction of those cells that are expressing CFP are stochastically simulated from the distributions described in Chapter 4. The same decision-

making algorithm as the actual Sortostat automation platform is used to decide if the cells in the sorting chamber should be sent to the waste or remain in the reactor. When the model simulates a cleaning event, the number of cells and the fraction of cells expressing CFP are also simulated stochastically from similar distributions as those used for the sorting event, however the cells are always sent to the waste. Thus, cleaning events only serve to increase the noise in the system and don't alter the direction of sorting preferentially.

I have run this simple model using initial conditions equivalent to one of the experimental runs. Sorting events and cleaning events occur at the same rates as in the experimental run, however in between those events nothing occurs in the model as I do not directly model cell growth. Even with such gross simplifications the model does a relatively good job of capturing the performance of the Sortostat (Figure 5-1).

Further work is needed to more appropriately account for growth in the Sortostat between sorting and cleaning events. Additionally since the model is stochastic it should be run many times to generate statistics on Sortostat performance. The results of those future simulations should be analyzed to provide metrics and performance limits for the strength of selective pressure that can be applied using the Sortostat.

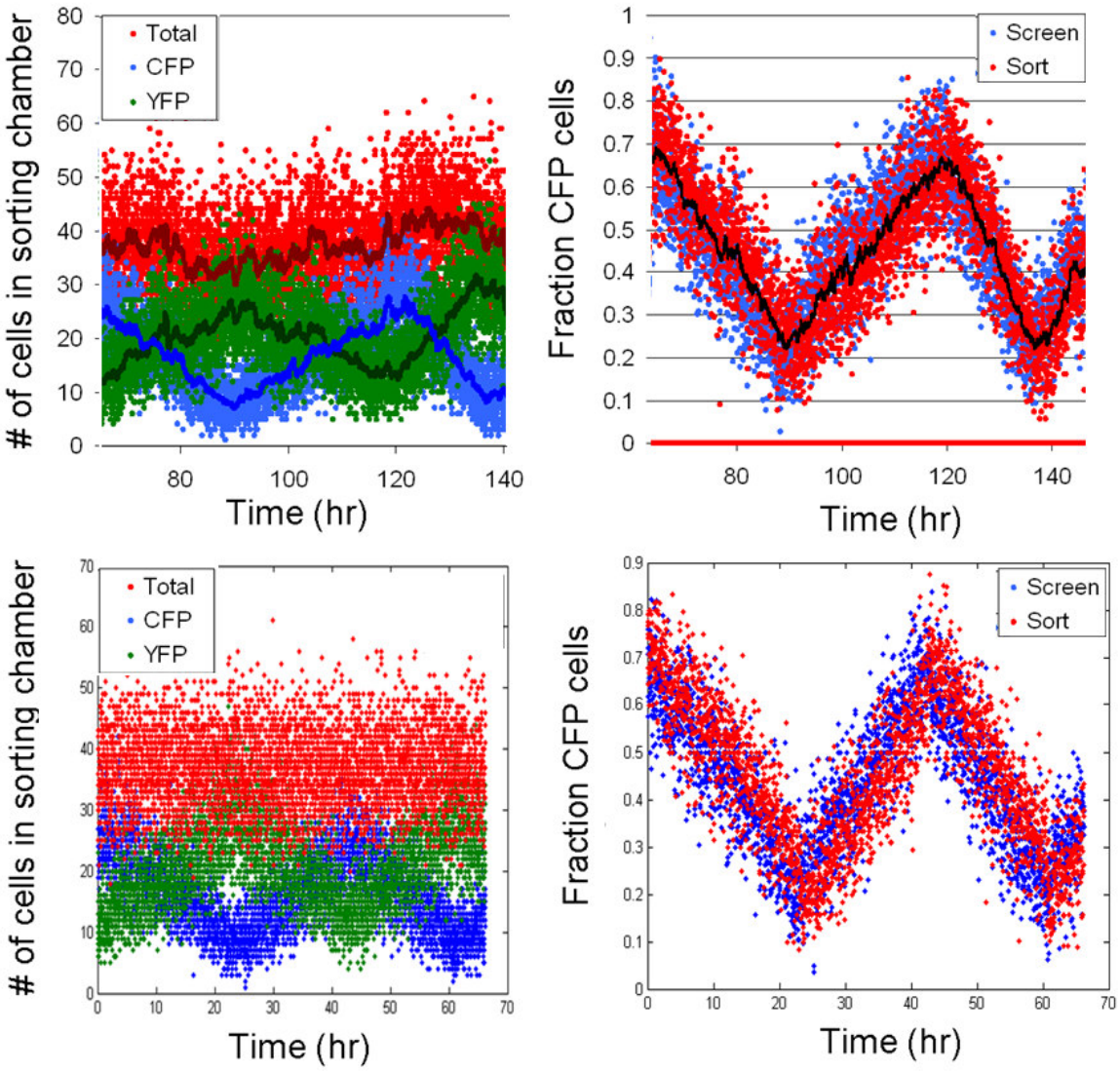


Figure 5-2 Comparison of Sortostat model with experimental results.

The two plots in the upper row shows a subsection of the experimental run described previously (Figure 4-12) and the two plots in the lower row shows one run of the stochastic model for Sortostat operation with initial conditions matching those in the experimental run. Since it is a stochastic model, the model output will vary each time the model is run, however these results are representative. The two plots in the left hand column shows time courses of the number of cells in the sorting chamber. For each time point there are three values shown: a blue dot represents the number of cells expressing CFP in the sorting chamber, a green dot represents the number of cells expressing YFP, and a red dot represents the total number of cells (CFP+YFP). The dark lines represent the 15 point moving average

for each measurement. The two plots in the right-hand column show the percentage of CFP expressing cells in the sorting chamber for each time point. The color of the points represents whether that particular sorting chamber was sent to the waste (Sorting event, red) or released back into the growth chamber loop (Screening event, blue). The black line represents the 15 point moving average.

Appendix A. Materials and Methods for Chapter 1

A.1 Bacterial Strains, Media, and Chemicals

E. coli strain MC4100 was used for all cloning steps. Cultures were grown in LB supplemented with antibiotics as needed. The *E. coli* strain CW2553 [98] carrying plasmid pJat8*araE* [32] was used for all characterization experiments. E. coli strain CW2553 (*araE201* Δ*araFGH*::*kan*) is a K-12 derivative with the wild-type genes for arabinose transport knocked out or mutated. In each experiment, I co-transformed CW2553 with pJat18 and pSB1A10. Supplemented M9 media (1x M9 salts, 1mM thiamine hydrochloride, 0.2% CAS amino acids, 2mM $MgSO_4$, 0.1 mM $CaCl_2$, carbon as appropriate) was used for characterization cultures. Ampicillin (at 100 μg/mL), kanamycin (50 μg/mL), gentamycin (20 μg/mL), and arabinose (various concentrations) were purchased from Sigma (St. Louis, MO). Restriction enzymes, T4 DNA ligase, and Taq and Phusion DNA polymerases were obtained from New England Biolabs (Ipswich, MA). All BioBrick parts and devices were obtained from the Registry of Standard Biological Parts [31].

A.2 Characterization of BioBrick Inverters and Terminators

Cells were grown overnight at 37C in 5ml of supplemented M9 minimal media (0.1% cas amino acids, 0.1% thiamine, and 0.4% glycerol) and antibiotics ampicillin and gentamycin. I diluted the cultures back 1 in 10 into supplemented M9 media. After 3 hours, the OD of the culture was taken and the cells were diluted to an OD of 0.0001 in fresh

supplemented M9 with antibiotics and arabinose concentrations 0%, 10^{-6}%, $3x10^{-6}$%, 10^{-5}%, $3x10^{-5}$%, 10^{-4}% w/w. Cells were placed at 37C and samples were collected after 14 hours following induction with arabinose. GFP and mRFP1 fluorescence levels for individual cells were measured on MOFLO flow cytometer. [http://dako.com] Cells were excited with a 488nm and 531nm laser and emission was collected though a GFP filter (530nm/30) and an RFP filter (650nm/LP), respectively. Fluorescence intensities were calibrated against beads [http://www.spherotech.com] with known intensities to account for day-to-day machine variation. The same protocol was used for the measurement of the two constructs in the RNAseE efficacy experiments (Figure 2-3).

A.3 Construction of inverter libraries

I constructed the inverter libraries by amplifying the BBa_Q04400 inverter using the VF and VR standard BioBrick primers and the GeneMorph II Random Mutagenesis Kit (Stratagene). I then digested and inserted the library into pSB1A10 using BioBrick standard assembly [29]. I transformed the assembled plasmids into electrocompetent CW2553 containing the plasmid pJat8*araE* plasmid. After transformation the cells were allowed to recover for 1 hr at 37ºC in LB with no antibiotic. The cells were then spun down and resuspended in 25 ml LB with ampicillin and gentamycin antibiotics and grown overnight at 37ºC to select for transformants. The overnight culture was divided into 1mL glycerols and stored at -80ºC. One of these stocks was thawed to conduct the two rounds of screening described in the Chapter 3 that led to the isolation of the functional mutant, BBa_Q04401.

A.4 Inverters in series

In order to compare the transfer curves of the two inverters BBa_Q04401 and BBa_Q04740 it was necessary to convert the input and output into the same units. I used a control version of pSB1A10 without any part or device inserted in order to establish the relationship between GFP and RFP fluorescence when PoPS input equals PoPS output (Figure A-1). This relationship can then be used to convert GFP fluorescence in relative arbitrary units into RFP fluorescence in relative arbitrary units. Once the input and output are in the same units the inverter transfer curves can be compared to predict whether the inverters in series will function properly. The relationship between GFP and RFP fluorescence expressed by the control pSB1A10 will be dependent on growth rate, among other factors. Since I know the growth rate was reduced significantly following insertion of the two inverters in series, it is not surprising that my predictions here were inaccurate

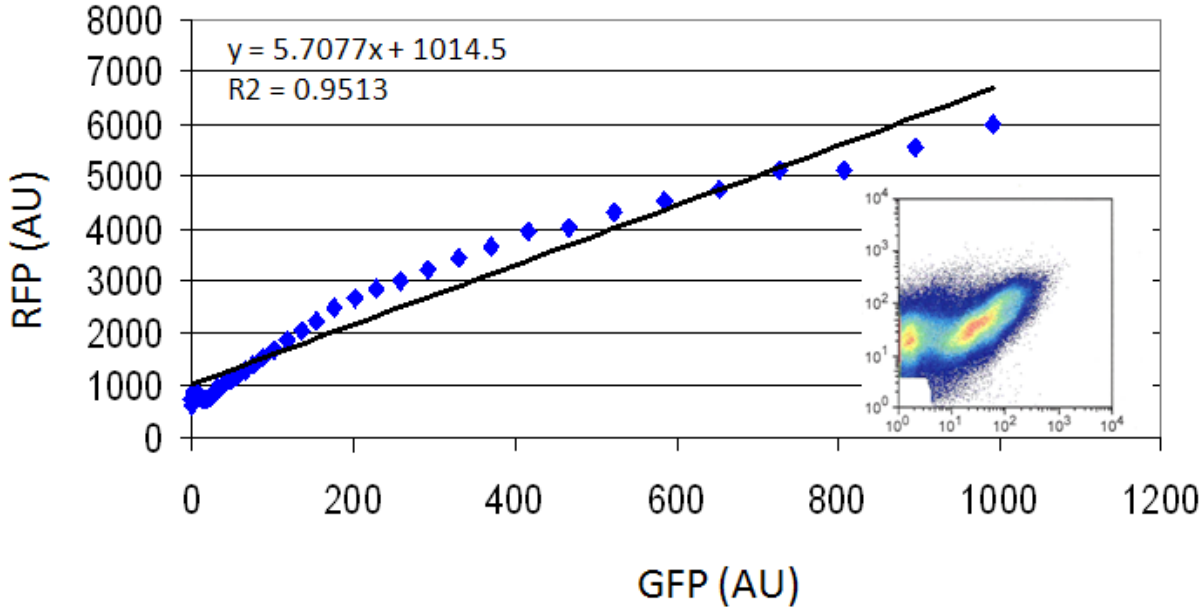


Figure. A-1 Relationship between GFP and RFP fluorescence measured by the same procedure used to characterize the inverters (Chapter A.2).

This relationship can be used to convert between GFP and RFP fluorescence in relative arbitrary units (AU) with the equation shown on the plot.

Appendix B. Materials and Methods for Chapter 2

B.1 Strains and Media

All measurement experiments and cloning were performed in *E.coli* TOP10 (Invitrogen). Supplemented M9 minimal medium (M9 salts, 1mM thiamine hydrochloride, 0.2% casamino acids, 0.1 M $MgSO_4$, 0.5 M $CaCl_2$) was used for all measurement experiments with either glycerol (0.4%) or glucose (0.4%) added as a carbon source and kanamycin (20 µg/ml) antibiotic added where appropriate. All oligonucleotides were purchased from Invitrogen and DNA modifying enzymes were purchased from New England Biolabs.

B.2 Kit contents

Sequences for all BioBrick plasmids (denoted pSB***) and BioBrick parts (denoted BBa_*****) are available through the Registry of Standard Biological Parts [45]. Physical copies of the plasmids and parts are also available from the Registry via the annual Registry parts distribution. The details of the promoter measurement kit contents are described in Box 1 and Table 1 and the details of the RBS measurement kit contents are described in Supplementary Box 1 and Table 1. The sequences for the preparative primers used to amplify pSB3K3 to generate backbone plasmid are: TACTAGTAGCGGCCGCTGCAG (forward primer) and CTCTAGAAGCGGCCGCGAATTC (reverse primer).

B.3 Construction of test constructs

We built promoters and RBSs by annealing synthesized oligonucleotides. The oligonucleotides were ordered with 5' phosphates and designed to leave an EcoRI overhang on the 5' end and a SpeI overhand on the 3' end so they could be used in subsequent ligation reactions without an intermediate restriction digest step. We inserted seven promoters: BBa_J23113, BBa_J23116, BBa_J23150, BBa_J23151, BBa_J23102, BBa_R0040, and BBa_R0011 into the promoter test construct and transformed into TOP10 according to the process outlined in BOX1. We inserted five RBSs: BBa_B0033, BBa_J61100, BBa_B0029, BBa_B0034, and BBa_B0035 into the RBS test construct and transformed into TOP10 according to the process outlined in Supplementary Box 1. We found the optimal concentration of DNA for each of the three components in the ligation reaction (pSB3K3, BBa_E0240 or BBa_I13401, and the test promoter or RBS) was approximately 10ng per uL. More detailed protocols and troubleshooting can be found at http://partsregistry.org/measurement. In the process of construction we found mutations in two of the promoters and one of the RBSs that we attribute to errors in the synthesis of the oligonucleotides that were annealed to construct the promoter and RBS. The two promoters and the RBS were functional, so we included them as additional members of the collection (BBa_J23150, BBa_J23151, and BBa_B0035). The method of part assembly described here is based on the three-antibiotic BioBrick standard assembly method [30].

B.4 Assay of promoter-RBS collection

The protocol described here will be referred to as the "original" protocol throughout the methods section and describes the measurement procedure used to characterize the set of seven promoters and five RBSs (Figure 3). For each promoter or RBS test construct three 17 mm test tubes containing 5 ml of pre-warmed (37°C) supplemented M9 medium with kanamycin (20 µg/ml) were inoculated from single colonies. Cultures were grown in 17 mm test tubes for approximately 20 hrs at 37°C with spinning at 70 rpm. We then diluted the cultures 1:100 into 5 ml of pre-warmed fresh media and the cultures were grown for approximately four hours under the previous conditions (17mm tubes, 37°C, spinning at 70 rpm). After four hours, we measured the OD600 of a 500 µl aliquot from each culture on a WPA Biowave Spectrophotometer. Based on this OD measurement, the cultures were diluted to the same OD (0.07) in 5 ml of pre-warmed fresh media and grown for one hour at 37°C. We then transferred three 200 µl aliquots from each culture into a flat-bottomed 96 well plate (Cellstar Uclear bottom, Greiner). We incubated the plate in a Wallac Victor3 multi-well fluorimeter (Perkin Elmer) at 37°C and assayed with an automatically repeating protocol of absorbance measurements (600 nm absorbance filter, 0.1 second counting time through 5 mm of fluid), fluorescence measurements (485 nm excitation filter, 525 nm emission filter, 0.1 seconds, CW lamp energy 12901 units), and shaking (3 mm, linear, normal speed, 15 seconds).

Background absorbance was determined by measuring wells containing only media. Background fluorescence was determined at different ODs from the fluorescence of TOP10

cells without a GFP expressing vector [99]. After background subtraction, time-series fluorescence (F) and absorbance (ABS) measurements were used to calculate the ratio of the rates of GFP synthesis for the promoter (or RBS) test construct and the reference standard construct. Measurements were taken from an approximately 30 min period in midexponential growth [100] (Supplementary Materials Figure 1 & 2). For example for the promoter measurement kit:

$$SPU = \frac{S^{SS}_{cell,\varphi}}{S^{SS}_{cell,J23101}} = \frac{(dF_{\varphi}/dt)/ABS_{\varphi}}{(dF_{\mathrm{J23101}}/dt)/ABS_{J23101}} \quad \text{(Eq. B.1)}$$

Since we are calculating a ratio of the GFP synthesis rates we do not need to determine each rate in absolute units of GFP per second per cell, rather we can use the background-subtracted fluorescence (F) that is proportional to the number of GFP molecules and the background-subtracted absorbance (ABS) that is proportional to the number of cells in the culture [49, 101] to calculate the ratio of GFP synthesis rates.

B.5 Assay of different measurement conditions

We measured the promoter activity of four promoters (BBa_J23113, BBa_J23150, BBa_J23151, and BBa_J23102) under four different measurement procedures. The first of the four procedures was identical to the "original" protocol described above for measuring the 12-member promoter and RBS collection. The second procedure was identical to the original except glucose was used as the carbon source in the media rather than glycerol. The third procedure was identical to the original except that instead of using the multi-well

fluorimeter to measure fluorescence and absorbance over time, we measured fluorescence on a flow cytometer (BD LSR II, Argon laser, Excitation: 488 nm, Emission filter: 530/30 nm) for a single time point (immediately after the 1hr incubation in culture tubes). The fourth procedure was identical to the original except that instead of using the multi-well fluorimeter to measure fluorescence and absorbance over time as the cells grew in the fluorimeter, we used the fluorimeter to measure only the first time point (immediately after the 1hr incubation in culture tubes).

For the first two procedures, we measured SPUs from the GFP synthesis rates as described in Equation B.1. For the last two measurement procedures we are unable to measure the GFP synthesis rates because these rates require a time series to calculate (dG/dt in Eq. B.1) and we collected only a single time point, however we can use this single time point to find the background-subtracted per cell fluorescence at steady-state ($[F]$). In the third procedure we use a flow cytometer to measure fluorescence per cell directly, thus $[F]$ is calculated by taking the geometric mean of the population fluorescence per cell. In the fourth procedure we use a fluorimeter and calculate fluorescence per cell by dividing fluorescence by absorbance ($[F] = F/ABS$). We related the per cell GFP concentration ($[G]$) to SPUs by using a model described previously [73] (derivation in Supplementary Materials):

$$SPU = \frac{[G]_{cell,\varphi}}{[G]_{cell,J23101}} * \frac{\mu_{\varphi}}{\mu_{J23101}} = \frac{[F]_{cell,\varphi}}{[F]_{cell,J23101}} * \frac{\mu_{\varphi}}{\mu_{J23101}} \quad \text{(Eq. B. 2)}$$

μ_φ/μ_{J23101} is a correction term based on differences in growth rate. Changes in the growth rate effect per cell GFP accumulation since loss of GFP per cell is largely due to dilution. Since we are calculating a ratio of the per cell GFP concentrations we do not need to determine each rate in absolute units of GFP molecules per cell, rather we can use the background-subtracted per cell fluorescence ($[F]$) that is proportional to the number of GFP molecules per cell to calculate the ratio of GFP concentrations.

B.6 Assay of inter-laboratory variability

We distributed a set of four promoters (BBa_J23113, BBa_J23150, BBa_J23151, and BBa_J23102) to six laboratories to take independent measurements of promoter activity. The protocol each lab conducted was identical to the original protocol described, except that the cells were harvested after the first 1:100 dilution and 4 hours of growth (there was no second dilution step). The cells were then spun down, resuspended in PBS, and the fluorescence per cell was measured using a flow cytometer. The measurement equipment used (cytometer model, laser, emission filter) varied between the laboratories (Supplementary Table 2). After background correction the per cell fluorescence ($[F]$) was determined for each promoter and activities in SPUs were calculated using Eq. B.2. We applied the growth rates measured previously (Supplementary Table 3) across all laboratories when calculating SPUs, rather than requesting individual laboratories to measure growth rates. This approximation likely increased the variability in the promoter activity

measurements across laboratories, as growth rates will vary between laboratories due to differences in culture conditions and media.

Appendix C. Materials and Methods for Chapter 3

C.1 Details of PDMS chip design and fabrication

The procedure for designing chips and requesting fabrication can be found on the website of the Kavli Nanoscience Institute Microfluidic Foundry at the California Institute of Technology [93]. I designed the Sortostat chip in AutoCAD 2004 (Autodesk Inc), and sent AutoCAD design files to CAD/Art Services (outputcity.com) who printed 20,000 dpi masks and sent these to the Foundry to be used for fabrication. The foundry proved to be an excellent source of PDMS chips, and allowed me to focus on design and operation of the Sortostat without needing to learn how to fabricate PDMS chips. There were some challenges in fabrication and I kept a log of the more common errors [102].

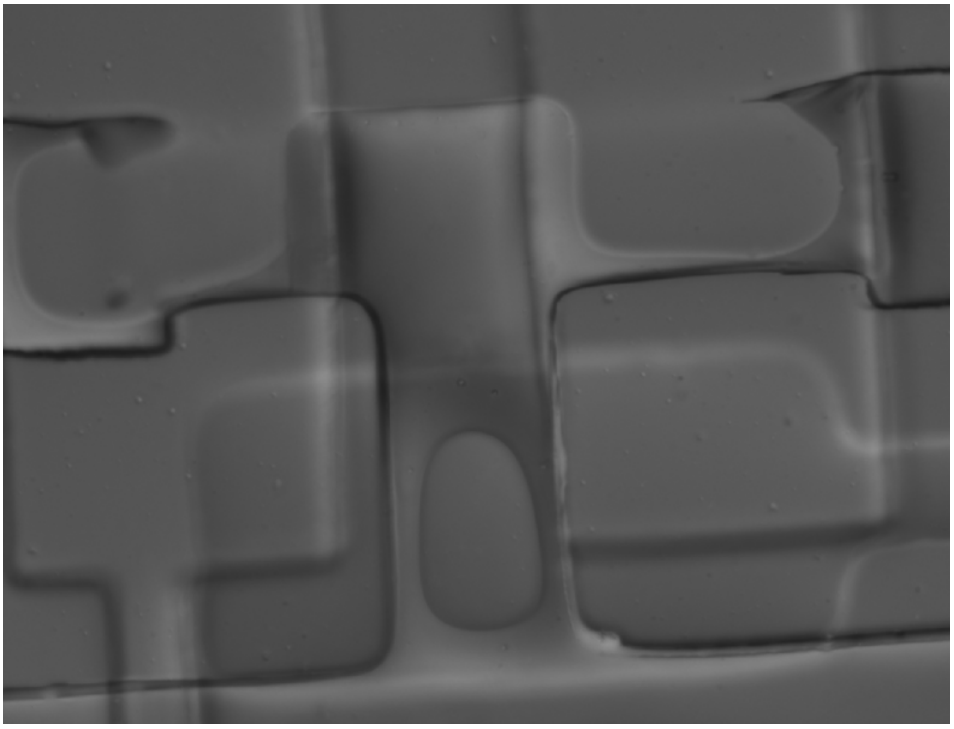


Figure. C-1 An example of a common fabrication error: valves that don't completely seal.

This fabrication error can come about as a result of an incorrect formulation of PDMS.

C.2 Previous designs of the Sortostat

My initial design for the Sortostat PDMS microfluidic chip included six Sortostat reactors per chip, operating in parallel. However, since the control lines are connected between all the reactors, a single fabrication error in a control line would often result in 6 non-functional Sortostats. Additionally, in experimental runs longer than one week I would often see a drop in cell growth rate that eventually led to washout of the culture. The cells were abnormally small and eventually would not grow in the presence of fresh media in the growth chamber loop. However, cells in the exit channels near one of the waste outlet ports

would grow vigorously; leading me to hypothesize that there was a problem with oxygen transfer through the PDMS to cells in the growth chamber loop. This may have been a result of the dense network of push-down valves (Figure C-2) present in this version of the design. These valves are pressurized at 1.7 atm and so might have created a pressure gradient that stopped the diffusion of oxygen into the chip. This hypothesis wasn't explored further after I redesigned the chip to only contain two Sortostat reactors. Cells growing in the new chips never displayed this small-cell phenotype.

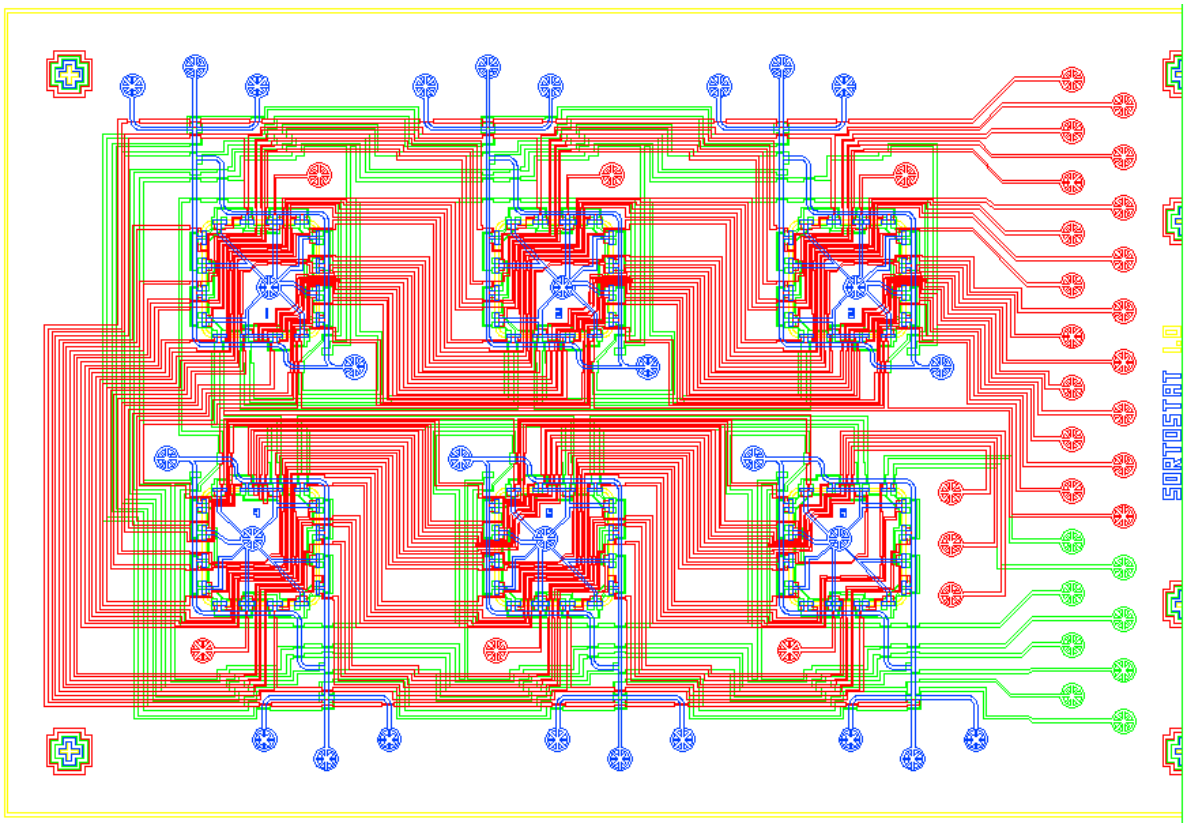


Figure. C-2 Schematic of the six reactor design of the Sortostat.

This design was replaced with the two reactor version due to issues with reliability (in both operation and fabrication) as well as possible problems with oxygen transfer through the dense network of push-down control channels (red).

C.3 Details of Sortostat automated platform

The Sortostat microfluidic chip contained 36 ports per reactor for connecting channels on the chip to off-chip pressure and media sources, 31 control channel ports and 6 fluid channel ports. I connected pressure and fluid sources to microfluidic channels using Tygon tubing (ID 0.020”, OD 0.060”; Cole-Parmer) and hollow steel pins (0.020” OD,

0.017" ID; New England Small Tube Corp) that connected to the ports. The tubing that was connected to control ports were filled with water, and prior to starting an experimental run I pressurize all control channels for ~30 minutes to allow the water in the tubing to displace the air in the channels. I also flowed media through the Sortostat fluid channels for 30 minutes in order to eliminate air in the channels as well as to coat the walls of the channels with Bovine Serum Albumin (BSA) that was present in the media. BSA reduces the amount of wall growth in the reactor. Media was introduced to the chip via a media bottle held under a constant pressure of ~4 psi. Control channels were connected to a ~20 psi pressure source via manifolds of 8 solenoid valves (Fluidigm) that was controlled by LabVIEW software (National Instruments). LabVIEW interfaced with the 8-valve manifolds via a digital input/output card (PCI-DIO-32HS) and a dedicated manifold controller (BOB-3 Microfluidic Valve Manifold Controller; Fluidigm). Prior to introducing cells into the chip, I grew the cells at 37°C to an OD600 of ~ 0.1 in supplemented M9 media (1x M9 salts, 1mM thiamine hydrochloride, 0.2% CAS amino acids, 2mM $MgSO_4$, 0.1 mM $CaCl_2$) with 0.4% glycerol and ampicillin (at 100 μg/mL).

Microscopy was performed using a Nikon TE2000 microscope and a Hamamatsu digital CCD camera (ORCA_AG C4742-80-12AG). Images were acquired using a 20x Plan Fluor DIC objective (Nikon). YFP was imaged using a YPF filter set (#41028; Chroma, Rockingham VT 05101). CFP was imaged using a CPF filter set (#31044; Chroma, Rockingham VT 05101). All LabView code necessary for operating the Sortostat is archived at MIT Dspace [94].

C.4 Comparison of the growth rate of CFP expressing and YFP expressing cells.

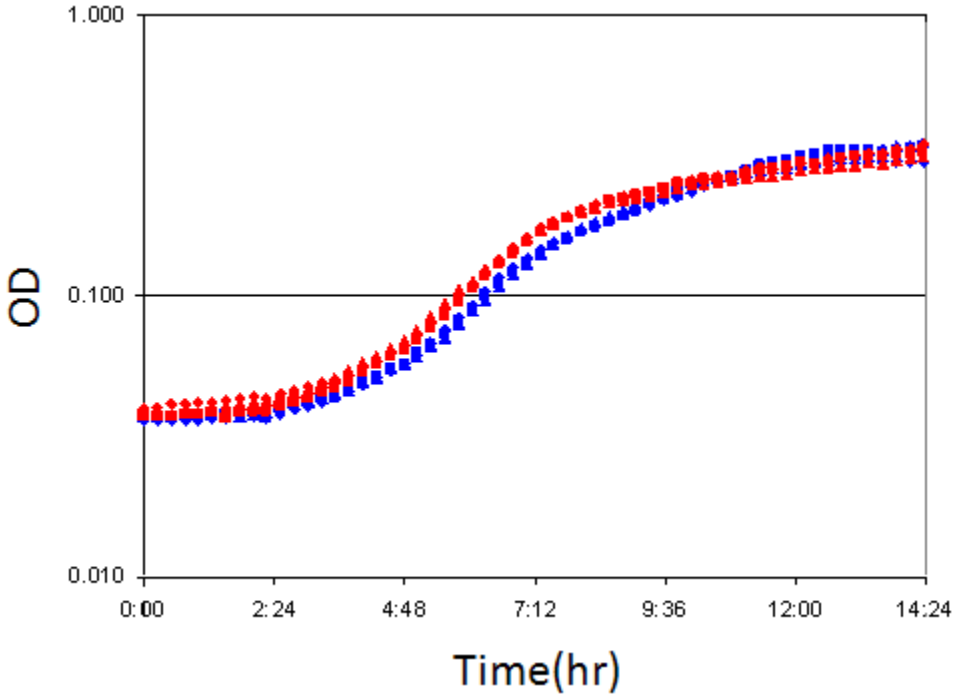


Figure. C-3 Comparison of the growth curves of the CFP expressing cells (blue) and YFP expressing cells (red) used in all experiments with the Sortostat.

The growth rates are approximately identical for the two strains. Three replicates are shown.

C.5 Image processing

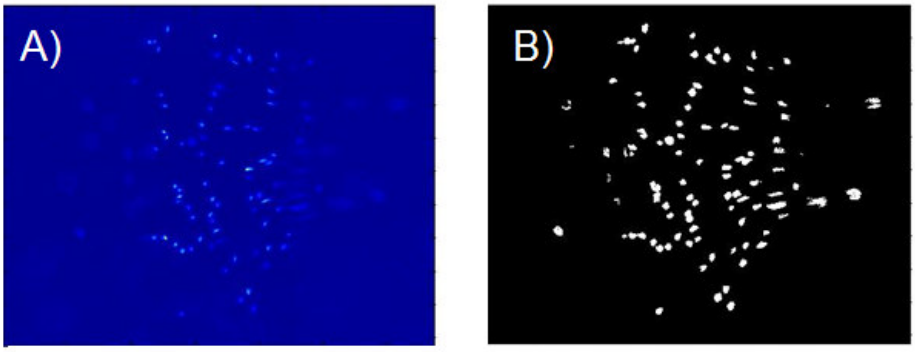


Figure. C-4 Example of image processing.

(A) unprocessed micrograph of CFP expressing cells in the sorting chamber. (B) Processed image white dots represent cells, the algorithm then counts groups of connected dots with more than 20 pixels as a cell and reports the total number of cells.

C.6 Thin channels for improved image processing

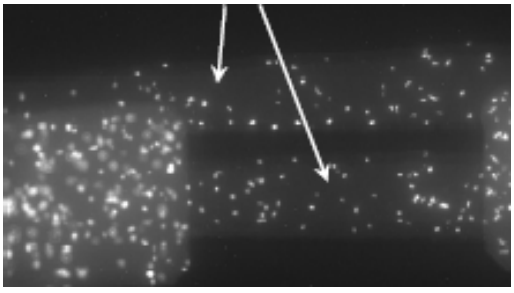


Figure. C-5 Micrograph of 3 µm high channels (arrows) that are used to bring the cells in a microchannel into a narrowed plane of focus. Image courtesy of Frederick Balagadde [83].

These thin channels could be added to a future version of the Sortostat in order to improve the quality of the images captured and allow for more complicated screens such as for cellular morphology.

C.7 Common failure modes

The first version of the Sortostat described here has about a 30% rate of successful operation. The most common failure mode is defects in chip fabrication that cause breaks in the control lines leading to dysfunctional operation of the device. This challenge could best be addressed by simplifying the overall chip design, possibly by reducing the number of cleaning chambers or by developing an effective passive method for preventing wall growth such as better surface coatings or a mutant strain with less likelihood of wall growth.

Another mode of failure is issues with the pressurized macro-fluidic lines that are used to control valves on the chip. On rare occasions the pins that connect these lines to the chip can pop out of the chip causing depressurization of a control line and often device failure. This failure mode could be countered by using a standard "pin-out" for microfluidic chips such that multiple pins were attached to the chip in a single unit. This would help to stabilize the junction between the pins and the chip and also reduce the setup time since pins would not need to be inserted individually.

Another common failure mode is incomplete sealing of the valves on the chip. Differences in the PDMS mixture used for chip fabrication can lead to chips with different elasticity that require different pressures for proper valve sealing. Failures due to valve sealing can largely be overcome by verifying the function of each valve by opening and closing the valve and observing function on the microscope before the runs starts. If valves are not sealing completely the pressure in the control lines can be increased until the valves

are functioning properly. However, at pressures greater than 30 psi the likelihood of a failure in the macrofluidic lines due to a pin popping out of the chip increases significantly.

Finally, about 10% of the failures were cases where the cells slowly died out in the reactor and I was unable to determine the source of the problem. I suspect that in many cases this cell death could have been due to incomplete valve sealing. Since some of the valves are only used infrequently, it would be challenging to diagnose if a valve stopped sealing well in the middle of a run and was leaking cells from the reactor out the waste. This failure mode could also be due to as-yet-unknown aspects of the growth of cells in the microfluidic environment.

Appendix D. Supplementary materials for Chapter 2

D.1 Modeling

D.1.1 Derivation of the relationships between absolute promoter or RBS activity and GFP synthesis rate or GFP concentration

In order to calculate the absolute promoter activity ($PoPS$) and RBS activity (ρ) from either the per cell GFP synthesis rate or the per cell GFP concentration at steady state for a promoter measured using the measurement kit, we extended a previously described ODE model [73] for GFP expression from a constitutive promoter. Our model is shown here:

$$\frac{d[M]}{dt} = n * PoPS - \gamma_M[M] \qquad (\text{Eq. } D.1)$$

$$\frac{d[I]}{dt} = \rho[M] - (a + \gamma_I)[I] \qquad (\text{Eq. } D.2)$$

$$\frac{d[G]}{dt} = a[I] - (\gamma_G)[G] \qquad (\text{Eq. } D.3)$$

We also define the per cell mature GFP synthesis term specifically as:

$$S_{cell} = a[I] \qquad (\text{Eq. } D.4)$$

Where $[M]$ is the concentration of mRNA, $[I]$ is the concentration of immature GFP, $[G]$ is the concentration of mature GFP, γ_M is the mRNA degradation rate, a is the GFP maturation rate, γ_I is the degradation rate of immature GFP, n is the copy number of the plasmid containing the promoter, ρ is the rate of synthesis of immature GFP in absolute units of

protein per second per mRNA, and $PoPS$ is the rate of mRNA synthesis in absolute units of successful mRNA initiation events per second per DNA copy of the promoter.

These equations can be used to establish four relationships where promoter and RBS activity ($PoPS$ and ρ, respectively) are specified as functions of the experimentally measurable terms in the model (GFP synthesis rate, or S_{cell} and GFP concentration, or $[G]$). Each of these relationships can be determined easily by assuming the system is at steady state, and thus $d[M]/dt = d[I]/dt = d[G]/dt = 0$. The results of the four cases are:

$$PoPS^{SS} as\ a\ f(S_{cell}^{SS}): \qquad PoPS^{SS} = \frac{\gamma_M(a+\gamma_I)S_{cell}^{SS}}{\rho an} \qquad (\text{Eq. } D.5)$$

$$PoPS^{SS} as\ a\ f([G]^{SS}): \qquad PoPS^{SS} = \frac{\gamma_M(a+\gamma_I)\gamma_G[G]_{cell}^{SS}}{\rho an} \qquad (\text{Eq. } D.6)$$

$$\rho^{SS}\ as\ a\ f(S_{cell}^{SS}): \qquad \rho^{SS} = \frac{\gamma_M(a+\gamma_I)S_{cell}^{SS}}{PoPSan} \qquad (\text{Eq. } D.7)$$

$$\rho^{SS} as\ a\ f([G]^{SS}): \qquad \rho^{SS} = \frac{\gamma_M(a+\gamma_I)\gamma_G[G]_{cell}^{SS}}{PoPSan} \qquad (\text{Eq. } D.8)$$

D.1.2 Derivation of promoter activity in SPUs as a function of GFP synthesis rate

As a reminder, we define SPUs as a relative measurement in terms of the activity of a test promoter φ and the activity of the reference standard promoter BBa_J23101 as:

$$Activity\ of\ Promoter\ \varphi\ (SPUs) = \frac{PoPS_{\varphi}^{SS}}{PoPS_{J23101}^{SS}} \qquad (\text{Eq. } D.9)$$

We derived promoter activity in SPUs as a function of GFP synthesis rate in the main text and will only review it briefly here. In the main text we use Eq. D.8 and D.5 along with a

number of assumptions of equivalent rates in the promoter test construct and the reference standard construct to derive the activity of promoter φ in SPUs as a function of the measured GFP synthesis rates (S_{cell}^{SS}):

$$Activity\ of\ Promoter\ \varphi\ (SPU) = \frac{S_{cell,\varphi}^{SS}}{S_{cell,J23101}^{SS}} \quad (\mathrm{Eq.}\ D.10)$$

D.1.3 Derivation of promoter activity in SPUs as a function of GFP concentration

We can derive promoter activity in SPUs as a function of per cell GFP concentration by combining Eq. D.9 and D.6 to yield:

$$Activity\ of\ Promoter\ \varphi\ (SPU) = \frac{\frac{\gamma_{M,\varphi}(a_\varphi + \gamma_{I,\varphi})\gamma_G[G]_{cell,\varphi}^{SS}}{\rho_\varphi a_\varphi n_\varphi}}{\frac{\gamma_{M,J23101}(a_{J23101} + \gamma_{I,J23101})\gamma_G[G]_{cell,J23101}^{SS}}{\rho_{J23101} a_{J23101} n_{J23101}}} \quad (Eq.\ D.11)$$

We again make the same assumptions described in the main text to equate several parameters for the promoter test construct and the reference standard construct, allowing for cancelling of terms. The only additional assumption that needs to be made is that mature GFP is stable so that protein degradation is negligible compared to dilution due to cellular growth ($\gamma_{G,\varphi} = \mu_\varphi$ and $\gamma_{G,J23101} = \mu_{J23101}$, where μ is the cellular growth rate). This assumption is reasonable as the GFP used here does not have a degradation tag. These assumptions allow us to simplify Eq. D.11 and find a simple relationship relating the activity of promoter φ in SPUs as a function of the measured GFP concentrations ($[G]_{cell}$), the growth rate of the cells

containing the promoter test construct (μ_{φ}), and the growth rate of the cells containing the reference standard (μ_{J23101}).

$$Activity\ of\ Promoter\ \varphi\ (SPU) = \frac{[G]_{cell,\varphi}}{[G]_{cell,J23101}} * \frac{\mu_{\varphi}}{\mu_{J23101}} \quad (\text{Eq. } D.12)$$

D.1.4 Derivation of RBS activity in SRUs as a function of GFP synthesis rate

Standard RBS Units (SRUs) are defined as the ratio of the steady-state activity of a user-specified RBS (φ) to the steady-state activity of the reference standard RBS (BBa_B0032):

$$Activity\ of\ RBS\ \varphi\ (SRUs) = \frac{\rho^{SS}_{RBS\ \varphi}}{\rho^{SS}_{B0032}} \quad (\text{Eq. } D.13)$$

where ρ^{SS} is the steady-state rate of protein synthesis in absolute units of proteins per second per mRNA and the subscripts φ or B0032 refer to the rates associated with a user-specified RBS (φ) or rates associated with the reference standard RBS (BBa_B0032), respectively. We can combine Eq. D.7 and D.13 and then eliminate some parameters by cancellation of terms:

$$Activity\ of\ RBS\ \varphi\ (SRU) = \frac{\frac{\gamma_{M,\varphi}(a_{\varphi} + \gamma_{I,\varphi})S^{SS}_{cell,\varphi}}{PoPS_{\varphi}a_{\varphi}n_{\varphi}}}{\frac{\gamma_{M,B0032}(a_{B0032} + \gamma_{I,B0032})S^{SS}_{cell,B0032}}{PoPS_{B0032}a_{B0032}n_{B0032}}} \quad (\text{Eq. } D.14)$$

We can make additional assumptions to further simplify Eq. D.14. First, we can assume that GFP expressed from either the test RBS construct φ or the reference standard RBS has an equivalent maturation rate ($a_{\varphi} = a_{B0032} = a$) since the two RBS are measured under the

same culture conditions. Second, since both constructs are present on the same backbone plasmid, we assume that they are present at the same copy number ($n_{\varphi} = n_{B0032}$). Third, since the promoters are identical in both constructs then they each produce the same mRNA sequence with the exception of the RBS sequence. We chose RBSs with similar lengths and sequences to minimize the differences in the mRNA, thus we assume that the mRNA degradation rates are equivalent ($\gamma_{M,\varphi} = \gamma_{M,B0032}$). mRNA degradation is also a function of dilution due to cellular growth, however the dilution rate is negligible relative to typical rates of active mRNA degradation in *E.coli* [76]. Since the promoters are identical we can also assume that the rate of successful mRNA initiation events per DNA copy of the promoter is equivalent ($PoPS_{\varphi} = PoPS_{B0032}$). Lastly, we can assume that immature GFP is stable so that protein degradation is negligible compared to dilution due to cellular growth ($\gamma_{I,\varphi} = \mu_{\varphi}$ and $\gamma_{I,B0032} = \mu_{B0032}$, where μ is the cellular growth rate). Following the above assumptions, Eq. D.14 can be simplified to:

$$Activity\ of\ RBS\ \varphi\ (SRU) = \frac{(a + \mu_{\varphi})S^{SS}_{cell,\varphi}}{(a + \mu_{\mathrm{B0032}})S^{SS}_{cell,\mathrm{B0032}}} \qquad (\mathrm{Eq.}\ D.15)$$

This equation can be simplified further, by noting that:

$$if\ \left|\mu_{\varphi} - \mu_{\mathrm{B0032}}\right| \ll a\ \ then\ \ \frac{(a + \mu_{\varphi})}{(a + \mu_{\mathrm{B0032}})} \approx 1 \qquad (\mathrm{Eq.}\ D.16)$$

Therefore, if we assume that the difference between the growth rate of cells containing the RBS test construct (μ_{φ}) and the growth rate of cells containing the reference standard construct (μ_{B0032}) is negligible compared to the maturation rate of GFP then Eq. D.15 can

be combined with Eq. D.16 yielding:

$$Activity\ of\ RBS\ \varphi\ (SRU) = \frac{S^{SS}_{cell,\varphi}}{S^{SS}_{cell,B0032}} \qquad (\text{Eq. } D.17)$$

D.1.5 Derivation of RBS activity in SRUs as a function of GFP concentration

We can derive RBS activity in SPUs as a function of per cell GFP concentration by combining Eq. D.8 and D.13 to yield:

$$Activity\ of\ RBS\ \varphi\ (SRU) = \frac{\frac{\gamma_{M,\varphi}(a_\varphi + \gamma_{I,\varphi})\gamma_G [G]^{SS}_{cell,\varphi}}{PoPS_\varphi a_\varphi n_\varphi}}{\frac{\gamma_{M,B0032}(a_{B0032} + \gamma_{I,B0032})\gamma_G [G]^{SS}_{cell,B0032}}{PoPS_{B0032} a_{B0032} n_{B0032}}} \qquad (\text{Eq. } D.18)$$

We again make the same assumptions described in the previous section to equate several parameters for the RBS test construct and the reference standard construct, allowing for cancelling of terms. The only additional assumption that needs to be made is that mature GFP is stable so that protein degradation is negligible compared to dilution due to cellular growth ($\gamma_{G,\varphi} = \mu_\varphi$ and $\gamma_{G,B0032} = \mu_{B0032}$, where μ is the cellular growth rate). This assumption is reasonable as the GFP used here does not have a degradation tag. These assumptions allow us to simplify Eq. D.18 and find a simple relationship relating the activity of RBS φ in SRUs as a function of the measured GFP concentrations ($[G]_{cell}$), the growth rate of the cells containing the RBS test construct (μ_φ), and the growth rate of the cells containing the reference standard (μ_{B0032}).

$$Activity\ of\ RBS\ \varphi\ (SRU) = \frac{[G]_{cell,\varphi}}{[G]_{cell,B0032}} * \frac{\mu_\varphi}{\mu_{B0032}} \qquad (\text{Eq. } D.19)$$

D.2 Alternate designs for the RBS and promoter measurement kits

To explore our design goals we considered an alternate design of the promoter measurement kit and the ribosome binding site measurement kit. The alternate promoter measurement kit was designed to allow promoters to be easily removed from the testing plasmid (for use in screening promoter libraries) by replacing the XbaI site on the GFP reporter device with a SpeI site. Thus, when a promoter was inserted it could be later removed by digesting with EcoRI and SpeI. However, this design was rejected because it formed a non-BioBrick standard junction (1bp change) between the promoter and RBS that was found to weaken the measured strength of some of the promoters compared to a standard BioBrick control (up to 50% reduction in expression, not shown).

The alternate RBS measurement kit design was optimized for simplifying the insertion of the RBS into the test plasmid. The promoter BBa_J23101 was included on the backbone plasmid upstream of the EcoRI site. A BioBrick RBS could then be digested with EcoRI and SpeI and inserted downstream of the promoter in one step, however inserting the RBS will generate a non-standard BioBrick junction between the promoter and RBS (the junction here would be nearly twice as long as the standard junction and contain an EcoRI and an XbaI site). We found that this non-standard junction led to loss of function (no GFP expressed, not shown) in some of the RBSs tested and thus we rejected the design.

D.3 Transcription start site prediction

We predicted the transcriptional start site for each of the promoters tested using the promoter measurement kit. The RBS measurement kit re-uses the same promoter (BBa_J23101) for all RBSs tested, so the transcriptional site is the same for all RBS test constructs. We predicted the transcriptional start site based on heuristics previously published compilations of *E.coli* sigma70 promoters [75]. The likely site of transcription initiation is within 1bp of the underlined A in each of the sequences below and the yellow highlighting represents the -10 region. All of the J23XXX sequences have identical sequences between the -10 region and the transcriptional start so we only show BBa_J23101 as a representative example of this set.

BBa_J23101:

tttacagctagctcagtcctaggtattatgctagctActagag

BBa_R0011:

aattgtgagcggataacaattgacattgtgagcggataacaagatactgagcacatActagag

BBa_R0040:

tccctatcagtgatagagattgacatccctatcagtgatagagatactgagcactActagag

D.4 Calculating the relationship between PoPS and SPUs

Researchers will often need to convert from SPUs to absolute units as absolute units tie promoter activity to the larger system of related reactions within the cell. 1 SPU is equivalent to the steady-state rate of successful mRNA initiation events per DNA copy of the

reference standard promoter BBa_J23101 ($PoPS^{SS}_{J23101}$). We can use the equation D.5 to estimate the activity of BBa_J23101 in PoPS based on the steady-state GFP synthesis rate of cells containing the reference standard construct.

$$PoPS^{SS} = \frac{\gamma_M(a+\gamma_I)S^{SS}_{cell}}{\rho a n} \quad (\text{Eq. } D.5)$$

In this case we will need to make estimates of all the rates that we were able to cancel out when making our relative measurement in SPUs (Results). These rates were previously parameterized with an identical reporter mRNA under similar conditions using identical equipment [49], and we will make use of those values here. The GFP maturation rate (a) was found to be 1.8E-3 sec^{-1}, the mRNA degradation rate (γ_M) was found to be 4.8E-3 sec^{-1}, and the translation rate from RBS BBa_B0032 was found to be 0.4 proteins sec^{-1} $mRNA^{-1}$. We also measured the growth rate to be 0.51 hr^{-1}, and we estimate the copy number of the reference standard plasmid to be approximately 40 copies per cell based on unpublished experiments in the Endy lab with a nearly identical construct (identical except that the promoter was BBa_R0011 rather than BBa_J23101) though the cells were grown under different conditions. However, constructs with the same origin of replication as the reference standard construct (p15A) have been shown to have copy numbers of 20-30[67] or as low as 14-16 [103]. We used the data described previously (in Results, Assay of promoter-RBS collection) of fluorescence and absorbance measured for cells in midexponential growth in a multi-well fluorimeter to estimate GFP synthesis rate:

$$S^{SS}_{cell} = dGFP/dt/CFU \quad (\text{Eq. } D.20)$$

We used the following conversions measured previously in the Endy lab [49] relating background subtracted absorbance (ABS) to colony forming units (CFU) and background subtracted fluorescence (G) to molecules of GFP (GFP).

$$CFU = 3.1E8 * ABS - 1.6E6 \quad (\text{Eq. } D.21)$$

$$GFP = 7.0E8 * G + 6.0E11 \quad (\text{Eq. } D.22)$$

Combining Eqs. D.20-22, and our measurements of fluorescence and absorbance we find GFP synthesis to be $S_{cell}^{SS} = 102.2$ GFP molecules cell^{-1} sec^{-1}. If we combine the rates described above ($\gamma_M, a, \gamma_I, \rho, n$) and this measure of GFP synthesis in Eq. D.5 we find:

$$1\ SPU \approx 0.03\ PoPS\ (mRNA\ sec^{-1} DNA\ copy\ of\ the\ promoter^{-1})$$

Since many of the rates are estimated from similar, but not identical systems (for instance, the plasmid copy number could vary by as much as 4-fold across previously described systems) we expect this relationship to be refined over time with better measurements of the rates associated with GFP expression from the reference standard construct.

D.5 Calculating the relationship between protein production rate and SRUs

Unpublished experiments in the Endy lab directly measuring protein number by Western blot and mRNA number by Northern blot have found the activity of the reference standard RBS BBa_B0032 to be 0.4 proteins per second per mRNA for a similar construct to the reference standard construct though under different growth conditions, thus:

$$1\ SRU = 0.4\ proteins\ sec^{-1} mRNA^{-1}$$

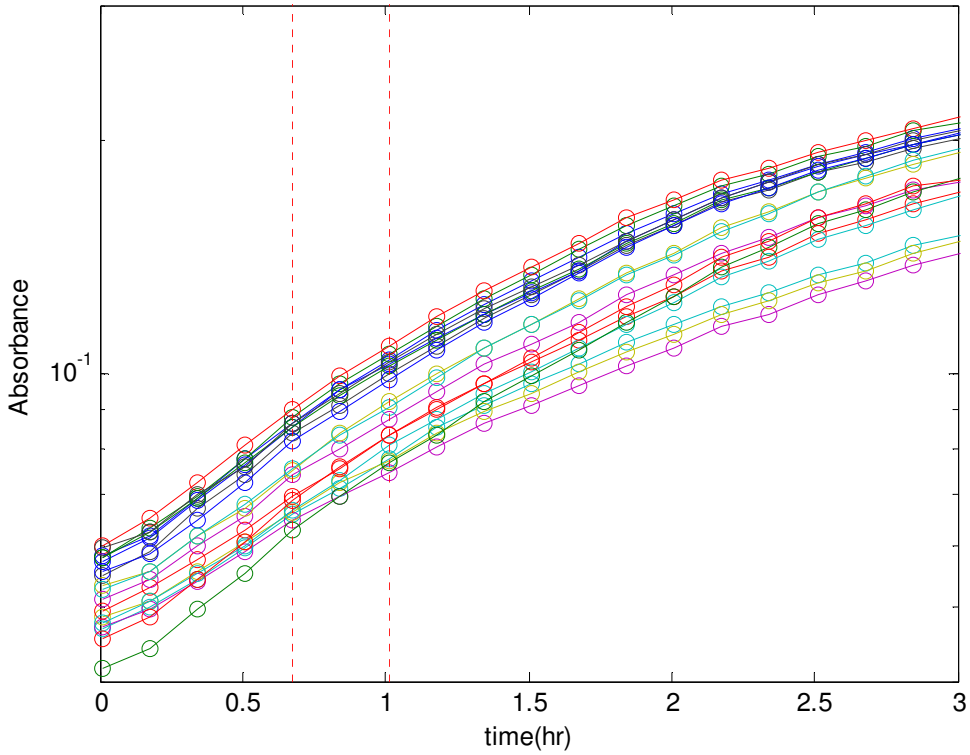


Figure. D-1 Example of typical growth curves

Growth curves for three replicates of each of the four promoters used in the inter-laboratory test (BBa_J23113, BBa_J23150, BBa_J23151, BBa_J23102), the reference standard promoter (BBa_J3101), and the negative control (TOP10 with no plasmid). The growth curves suggest that the cells are in exponential growth for about an hour (0.5-1.5 hrs), the vertical dotted lines indicate the region sampled to calculate the steady state GFP synthesis.

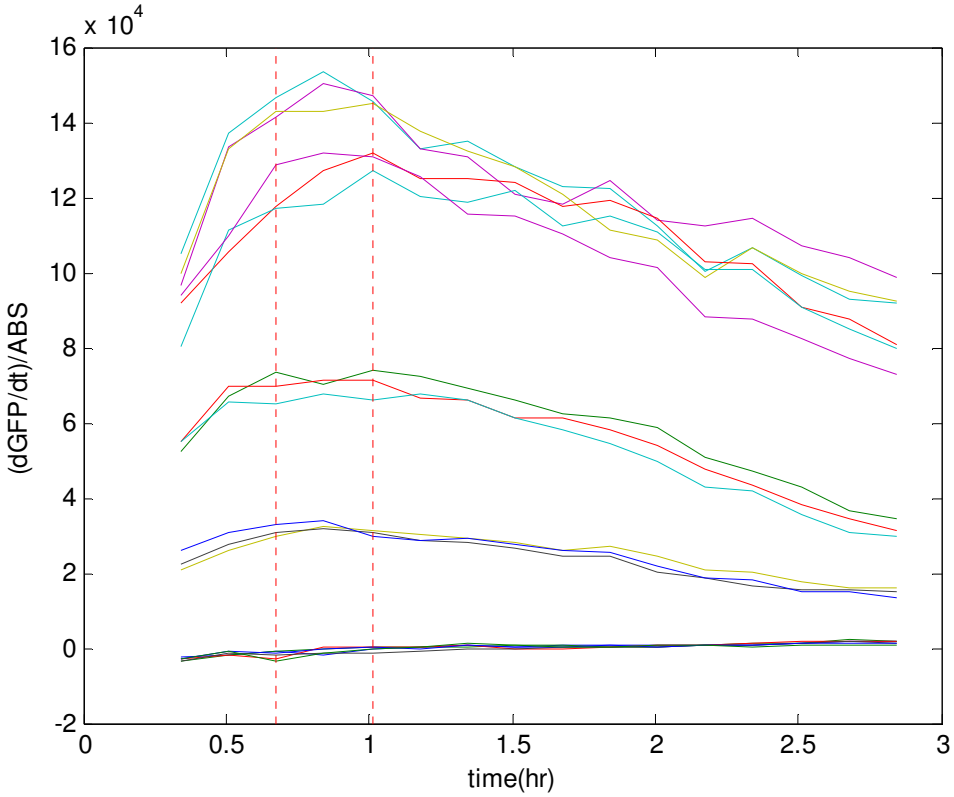


Figure. D-2 Example of GFP synthesis rate time series

The GFP synthesis rate (d[GFP]/dt / ABS) for three replicates of each of the four promoters used in the inter-laboratory test (BBa_J23113, BBa_J23150, BBa_J23151, BBa_J23102), the reference standard promoter (BBa_J23101), and the negative control (TOP10 with no plasmid). The order of the promoters from lowest to highest activity is: BBa_J23113, BBa_J23150, BBa_J23151, BBa_J23102, BBa_J23101. The activity of the BBa_J23113 promoter is not above the negative control within the limits of our measurement. There is a region of steady-state GFP synthesis from approximately 0.5 hrs to 1.25hrs. The vertical dotted lines indicate the region sampled to calculate the steady-state GFP synthesis rate.

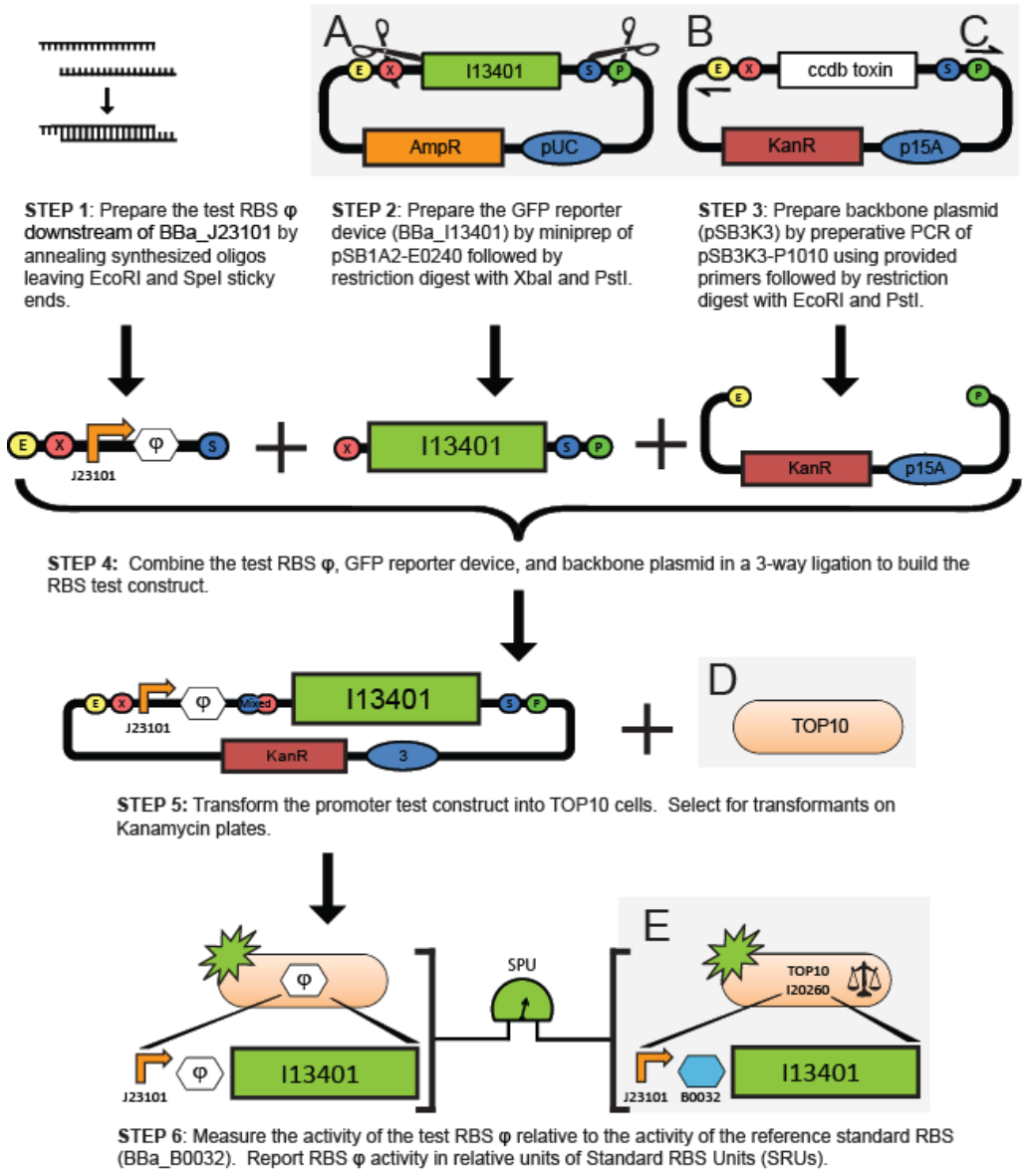


Figure. D-3 Instructions for inserting an RBS into the RBS measurement kit and measuring the promoter activity in Standard RBS Units (SRUs).

The components marked A-E and shaded in gray are included with the kit: **(A)** pSB1A2-I13401 is provided as purified DNA. BBa_I13401 is the GFP reporter device containing a GFP coding region (BBa_E004) and a transcriptional terminator (BBa_B0015). **(B)** pSB3K3-P1010 is provided as purified DNA to serve as template for a preparative PCR

reaction. **(C)** DNA primers are provided for use in a preparative PCR reaction to produce backbone vector (pSB3K3). **(D)** TOP10 cells are provided to serve as the standard characterization and construction strain. **(E)** pSB3K3-I20260 is provided as a plasmid in TOP10. BBa_I20260 contains the RBS reference standard, BBa_B0032, upstream of the GFP measurement device.

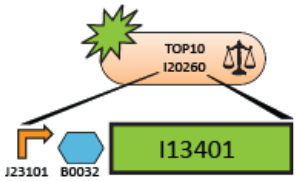
Notation	Name (BioBrick Part #)	Description / Function
	GFP reporter device (BBa_I13401)	The **GFP reporter device** contains the GFP coding region and a transcription terminator. It converts a translation input into molecules of GFP.
	Backbone plasmid (pSB3K3-P1010)	The **backbone plasmid** contains an expression cassette for ccdb toxin that serves as a counter-selectable marker as well as the p15A origin of replication and a kanamycin resistance gene.
	Test RBS φ	The **test RBS φ** is the user-specified RBS to be measured.
	RBS test construct	The **RBS test construct** contains the **test RBS** φ upstream of the **GFP reporter device** within the pSB3K3 **backbone plasmid.** The rate of GFP expressed from this construct is used to measure the activity of the test RBS.
	Reference standard construct (BBa_I20260)	The **reference standard construct** is identical to the **RBS test construct** except it contains the reference standard RBS (BBa_B0032). The rate of GFP expressed from this construct is used to measure the activity of the reference standard RBS.
	Preparative primers (BBa_G1000, BBa_G1001)	The **preparative primers** are used to amplify the pSB3K3 **backbone plasmid**.
	TOP10 (BBa_V1009)	*E. coli* TOP10 is used as the standard strain for kit measurement experiments.

Table. D-1 Components of RBS measurement kit

The nomenclature listed is used to describe the RBS measurement kit components within the text. The notation for each component serves as a key for the kit instructions (Supplementary Box1). More detailed information about each component such as the DNA sequence can be found at the Registry of Standard Biological Parts (http://partsregistry.org).

Make	Laser	Excitation Line	Emission Filter
BD FACSCalibur	Argon	488nm	530/30
BD LSR II	Argon	488nm	530/30
BD LSR II	Coherent Sapphire	488nm	530/30
Partec CyFlow Space	Solid state	488nm	520/30
BD FACSAria	Coherent Sapphire	488nm	535/30
BD LSR II	Argon	488nm	525/50
BD FACSAria	Coherent Sapphire	488nm	530/30

Table. D-2 Listing of the flow cytometer equipment used at the seven laboratories that participated in the inter-laboratory variability study.

	$\frac{\mu_\varphi}{\mu_{J23101}}$ (mean)	$\frac{\mu_\varphi}{\mu_{J23101}}$ (stdev)
TOP10	1.1063	0.1413
J23113	1.1264	0.104
J23150	1.21	0.1085
J23151	1.1875	0.1081
J23102	1.0985	0.0707

Table. D-3 Growth rates of cells containing the 4 promoter test constructs that were used in the inter-laboratory study measured relative to the reference standard.

Chapter 6. References

1. Kelly K: *Out of Control: The New Biology of Machines, Social Systems, & the Economic World.* Basic Books; 1995.
2. Farinas ET, Bulter T, Arnold FH: **Directed enzyme evolution.** *Current Opinion in Biotechnology* 2001, **12:**545-551.
3. Horwitz MS, Loeb LA: **DNA sequences of random origin as probes of Escherichia coli promoter architecture.** *The Journal of biological chemistry* 1988, **263:**14724-14731.
4. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY: **A monomeric red fluorescent protein.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99:**7877-7882.
5. Koza JR, Keane MA, Streeter MJ: *Genetic programming IV: Routine Human-Competitive Machine Intelligence.* Springer; 2003.
6. Spector L, Barnum H, Bernstein HJ, Swamy N: **Finding a better-than-classical quantum AND/OR algorithm using genetic programming.** In *Evolutionary Computation, 1999 CEC 99 Proceedings of the 1999 Congress on*. 1999: 2246 Vol. 2243-2246 Vol. 2243.
7. Adami C: *Introduction to Artificial Life.* Springer; 1998.
8. Winkler WC, Breaker RR: **Genetic control by metabolite-binding riboswitches.** *Chembiochem : a European journal of chemical biology* 2003, **4:**1024-1032.
9. Zhang J: **On the Evolution of Codon Volatility.** *Genetics* 2005, **169:**495-501.
10. Endy D: **Foundations for engineering biology.** *Nature* 2005, **438:**449-453.
11. Arnold FH, Volkov AA: **Directed evolution of biocatalysts.** *Current Opinion in Chemical Biology* 1999, **3:**54-59.
12. Kuchner O, Arnold FH: **Directed evolution of enzyme catalysts.** *Trends in Biotechnology* 1997, **15:**523-530.
13. Shetty R: **Engineering Transcription-based logic.** Massachusetts Institute of Technology, 2008.
14. Braff J: **Personal Communication (unpublished).** 2005.
15. Stemmer WP: **Rapid evolution of a protein in vitro by DNA shuffling.** *Nature* 1994, **370:**389-391.
16. Guet ClC, Elowitz MB, Hsing W, Leibler S: **Combinatorial synthesis of genetic networks.** *Science (New York, NY)* 2002, **296:**1466-1470.
17. Itaya M, Fujita K, Kuroki A, Tsuge K: **Bottom-up genome assembly using the Bacillus subtilis genome vector.** *Nat Meth* 2008, **5:**41-43.
18. Yokobayashi Y, Arnold F: **A Dual Selection Module for Directed Evolution of Genetic Circuits.** *Natural Computing* 2005, **4:**245-254.

19. Yokobayashi Y, Weiss R, Arnold FH: **Directed evolution of a genetic circuit.** *Proceedings of the National Academy of Sciences* 2002, **99:**16587-16591.
20. Anderson JC, Voigt CA, Arkin AP: **Environmental signal integration by a modular AND gate.** *Molecular Systems Biology* 2007, **3**.
21. Voigt CA: **Genetic parts to program bacteria.** *Current Opinion in Biotechnology* 2006, **17:**548-557.
22. Elowitz MB, Leibler S: **A synthetic oscillatory network of transcriptional regulators.** *Nature* 2000, **403:**335-338.
23. Gardner TS, Cantor CR, Collins JJ: **Construction of a genetic toggle switch in Escherichia coli.** *Nature* 2000, **403:**339-342.
24. Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD: **Engineering a mevalonate pathway in Escherichia coli for production of terpenoids.** *Nature biotechnology* 2003, **21:**796-802.
25. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C: **Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli.** *Science (New York, NY)* 2001, **291:**1790-1792.
26. Hasty J: **Design then mutate.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**.
27. Endy: **Adventures in synthetic biology.** *Nature* 2005.
28. Khlebnikov A, Skaug T, Keasling JD: **Modulation of gene expression from the arabinose-inducible araBAD promoter.** *Journal of industrial microbiology & biotechnology* 2002, **29:**34-37.
29. Knight T: **Idempotent Vector Design for Standard Assembly of Biobricks.** 1999.
30. Shetty R, Endy D, Knight T: **Engineering BioBrick vectors from BioBrick parts.** *Journal of Biological Engineering* 2008, **2:**5-5.
31. **Registry of Standard Biological Parts** [http://partsregistry.org]
32. Khlebnikov A, Datsenko KA, Skaug T, Wanner BL, Keasling JD: **Homogeneous expression of the PBAD promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter.** *Microbiology* 2001, **147:**3241-3247.
33. Smolke CD, Keasling JD: **Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon.** *Biotechnology and bioengineering* 2002, **80:**762-776.
34. Choe J, Guo HH, van den Engh G: **A dual-fluorescence reporter system for high-throughput clone characterization and selection by cell sorting.** *Nucl Acids Res* 2005, **33:**e49-e49.
35. Nicholson AW: **Function, mechanism and regulation of bacterial ribonucleases.** *FEMS microbiology reviews* 1999, **23:**371-390.
36. Andersen JB, Sternberg C, Poulsen LK, Bjorn SP, Givskov M, Molin S: **New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria.** *Applied and environmental microbiology* 1998, **64:**2240-2246.

37. Payne S, Shetty R, Venkatachalam V, Broadbent K, Green II D, Zhu B, Canton B, Che A, Kelly J, Sutton S, et al: **A synthetic biology approach to reprogramming bacterial odor.** *Submitted*.
38. Hooshangi S, Thiberge S, Weiss R: **Ultrasensitivity and noise propagation in a synthetic transcriptional cascade.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**.
39. Pedraza JM, van Oudenaarden A: **Noise Propagation in Gene Networks.** *Science* 2005, **307:**1965-1969.
40. Knight T: **Idempotent Vector Design for Standard Assembly of Biobricks.** 2003. doi:1721.1/21168
41. Arkin AP, Endy D: **A Standard Parts List for Biological Circuitry.** DARPA White Paper; 1999. doi:1721.1/29794
42. Endy D: **2003 Synthetic Biology study.** 2003. doi: 1721.1/38455
43. Knight T: **DARPA BioComp Plasmid Distribution 1.00 of Standard Biobrick Components.** 2002. doi:1721.1/21167
44. Texas Instruments: *TTL logic data book: standard TTL, Schottky, low-power Schottky.* Texas Instruments; 1988.
45. Spicer CC: **The Theory of Bacterial Constant Growth Apparatus.** *Biometrics* 1955, **11:**225-230.
46. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB: **Accurate prediction of gene feedback circuit behavior from component properties.** *Mol Syst Biol* 2007, **3**.
47. Ajo-Franklin CM, Drubin DA, Eskin JA, Gee EPS, Landgraf D, Phillips I, Silver PA: **Rational design of memory in eukaryotic cells.** *Genes Dev* 2007, **21:**2271-2276.
48. Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ: **A bottom-up approach to gene regulation.** *Nature* 2006, **439:**856-860.
49. Canton B, Labnoa A, Endy D: **BBa_F2620, an Engineered Cell-Cell Communication Receiver Device.** *Nature Biotechnology (in press)* 2008.
50. Serebriiskii IG, Golemis EA: **Uses of lacZ to Study Gene Function: Evaluation of [beta]-Galactosidase Assays Employed in the Yeast Two-Hybrid System.** *Analytical Biochemistry* 2000, **285:**1-15.
51. Miller JH: *Experiments in Molecular Genetics.* Cold Spring HarborLaboratory; 1972.
52. Eustice DC, Feldman PA, Colberg-Poley AM, Buckery RM, Neubauer RH: **A sensitive method for the detection of beta-galactosidase in transfected mammalian cells.** *BioTechniques* 1991, **11:**739-740, 742-733-739-740, 742-733.
53. Kawasaki ES: **The End of the Microarray Tower of Babel: Will Universal Standards Lead the Way?** *J Biomol Tech* 2006, **17:**200-206.
54. Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, Causton H, Conley MP, Elespuru R, Fero M, et al: **The External RNA Controls Consortium: a progress report.** *Nature methods* 2005, **2:**731-734.
55. Tong W, Lucas AB, Shippy R, Fan X, Fang H, Hong H, Orr MS, Chu T-M, Guo X, Collins PJ, et al: **Evaluation of external RNA controls for the assessment of microarray performance.** *Nat Biotech* 2006, **24:**1132-1139.

56. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, et al: **Minimum information about a microarray experiment (MIAME)[mdash]toward standards for microarray data.** *Nat Genet* 2001, **29:**365-371.
57. Kelvin WT, Joule JP, Maxwell JC: *Reports of the Committee on Electrical Standards Appointed by the British.* E. & F.N. Spon; 1873.
58. Hunt BJ: **The Ohm Is Where the Art Is: British Telegraph Engineers and the Development of Electrical Standards.** *Osiris* 1994, **9:**48-63.
59. Unger MA, Chou HP, Thorsen T, Scherer A, Quake SR: **Monolithic microfabricated valves and pumps by multilayer soft lithography.** *Science (New York, NY)* 2000, **288:**113-116.
60. Zarrinpar A, Park S-H, Lim WA: **Optimization of specificity in a cellular protein interaction network by negative selection.** *Nature* 2003, **426:**676-680.
61. Elowitz MB, Levine AJ, Siggia ED, Swain PS: **Stochastic Gene Expression in a Single Cell.** *Science* 2002, **297:**1183-1186.
62. Alper H, Fischer C, Nevoigt E, Stephanopoulos G: **Tuning genetic control through promoter engineering.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**.
63. Weiss R: **Cellular computation and communications using engineered genetic regulatory networks.** Massachusetts Institute of Technology, 2001.
64. Southward CM, Surette MG: **The dynamic microbe: green fluorescent protein brings bacteria to light.** *Molecular Microbiology* 2002, **45:**1191-1196.
65. Cormack BP, Valdivia RH, Falkow S: **FACS-optimized mutants of the green fluorescent protein (GFP).** *Gene* 1996, **173:**33-38.
66. Kobayashi H, Kaern M, Araki M, Chung K, Gardner TS, Cantor CR, Collins JJ: **Programmable cells: Interfacing natural and engineered gene networks.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**.
67. Lutz R, Bujard H: **Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements.** *Nucleic acids research* 1997, **25:**1203-1210.
68. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R: **A synthetic multicellular system for programmed pattern formation.** *Nature* 2005, **434:**1130-1134.
69. Horn GT, Wells RD: **The leftward promoter of bacteriophage lambda. Structure, biological activity, and influence by adjacent regions.** *J Biol Chem* 1981, **256:**2003-2009.
70. Brunner M, Bujard H: **Promoter recognition and promoter strength in the Escherichia coli system.** *The EMBO Journal* 1987, **6**.
71. Rao L, Ross W, Appleman JA, Gaal T, Leirmo S, Schlax PJ, Record MT, Gourse RL: **Factor independent activation of rrnB P1. An "extended" promoter with an upstream element that dramatically increases promoter strength.** *Journal of Molecular Biology* 1994, **235:**1421-1435.

72. Endy: **Adventures in synthetic biology.** *Nature* 2005, **438:**449 - 453.
73. Leveau JHJ, Lindow SE: **Predictive and Interpretive Simulation of Green Fluorescent Protein Expression in Reporter Bacteria.** *Journal of Bacteriology* 2001, **183**.
74. Adams CW, Hatfield GW: **Effects of promoter strengths and growth conditions on copy number of transcription-fusion vectors.** *J Biol Chem* 1984, **259:**7399-7403.
75. Hawley DK, McClure WR: **Compilation and analysis of Escherichia coli promoter DNA sequences.** *Nucleic acids research* 1983, **11:**2237-2255.
76. Bernstein JA, Khodursky AB, Lin P-H, Lin-Chao S, Cohen SN: **Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays.** *Proceedings of the National Academy of Sciences* 2002, **99:**9697-9702.
77. Bagh S, Mazumder M, Velauthapillai T, Sardana V, Dong GQ, Movva AB, Lim LH, McMillen DR: **Plasmid-borne prokaryotic gene expression: Sources of variability and quantitative system characterization.** *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)* 2008, **77:**021919-021912.
78. Bremer H, Dennis P: **Modulation of chemical composition and other parameters of the cell by growth rate.** In *Escherichia coli and Salmonella 2nd edn. Volume* 2. Edited by Neidhart FC. Washington DC: ASM Press; 1996: 1553–1569
79. Bradbury J: **Nature's Nanotechnologists: Unveiling the Secrets of Diatoms.** *PLoS Biology* 2004, **2:**e306 EP --e306 EP -.
80. Gordon R: **Computer controlled evolution of diatoms: design for a compustat.** *Nova Hedwigia* 1996, **112:**213-216.
81. Thomson T: **Tools for Model Building and Time-Dependent Investigation of Signaling in the Yeast Mating Response System.** Massachusetts Institute of Technology, Biological Engineering; 2008.
82. Balagadde FK, You L, Hansen CL, Arnold FH, Quake SR: **Long-term monitoring of bacteria undergoing programmed population control in a microchemostat.** *Science (New York, NY)* 2005, **309:**137-140.
83. Balagaddé FK: **Description of the microchemostat.** California Institute of Technology, 2007.
84. Herbert D, Elsworth R, Telling RC: **The continuous culture of bacteria; a theoretical and experimental study.** *Journal of general microbiology* 1956, **14:**601-622.
85. Novick A, Szilard L: **Description of the chemostat.** *Science (New York, NY)* 1950, **112:**715-716.
86. **Chemostat** [http://en.wikipedia.org/wiki/Chemostat]
87. Wiebe MG, Robson GD, Shuster J, Trinci AP: **Evolution of a recombinant (gucoamylase-producing) strain of Fusarium venenatum A3/5 in chemostat culture.** *Biotechnology and bioengineering* 2001, **73:**146-156.

88. Jannasch HW, Mateles RI: **Experimental bacterial ecology studied in continuous culture.** In *Advances in Microbial Physiology.* Edited by Rose AH, Tempest DW. London: Academic Press; 1974: 165-212
89. Smith HL, Waltman P: *The Theory of the Chemostat: Dynamics of Microbial Competition.* Cambridge UniversityPress; 1995.
90. Costerton JW, Lewandowski Z, Caldwell DE, Korber DR, Lappin-Scott HM: **Microbial biofilms.** *Annual review of microbiology* 1995, **49:**711-745.
91. Topiwala HH, Hamer C: **Effect of Wall Growth in Steady-State Continuous Culture.** *Biotechnology and Bioengineering* 1971, **13:**919-922.
92. Hansen CL: California Institue of Technology, Applied Physics; 2004.
93. **Kavli Nanoscience Institute Microfluidic Foundry at the California Institute of Technology** [http://kni.caltech.edu/foundry/index.html]
94. **LabView VIs for operating Sortostat** [http://hdl.handle.net/1721.1/41533]
95. Bryson V, Szybalski W: **Microbial selection.** *Science (New York, NY)* 1952, **115:**45-51.
96. Gralla JD, Collado-Vides J: **Organization and Function of Transcription Regulatory Elements.** In *Cellular and Molecular Biology: Escherichia coli and Salmonella.* Edited by F.C. N. Washington, D.C.: American Society for Microbiology; 1996: 1232-1245
97. Hsu LM: **Promoter Escape by Escherichia coli RNA Polymerase.** In *Cellular and Molecular Biology: Escherichia coli and Salmonella.* Edited by F.C. N. Washington, D.C.: American Society for Microbiology; 2008
98. Horazdovsky BF, Hogg RW: **Genetic reconstitution of the high-affinity L-arabinose transport system.** *Journal of Bacteriology* 1989, **171**.
99. Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, Leibler S, Surette MG, Alon U: **Ordering Genes in a Flagella Pathway by Analysis of Expression Kinetics from Living Bacteria.** *Science* 2001, **292:**2080-2083.
100. Setty Y, Mayo AE, Surette MG, Alon U: **Detailed map of a cis-regulatory input function.** *Proceedings of the National Academy of Sciences* 2003, **100:**7702-7707.
101. Ronen M, Rosenberg R, Shraiman BI, Alon U: **Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics.** *Proceedings of the National Academy of Sciences* 2002, **99:**10555-10560.
102. **Foundry Report Card** [http://openwetware.org/index.php?title=Sortostat/Foundry_report_card&oldid=179236]
103. Hiszczyska-Sawicka E, Kur J: **Effect of Escherichia coli IHF mutations on plasmid p15A copy number.** *Plasmid* 1997, **38:**174-179.